Vehicle Distributions in Large and Small Cities: Spatial Models and Applications

Qimei Cui, *Senior Member, IEEE,* Ning Wang, and Martin Haenggi, *Fellow, IEEE*

***Abstract*—Safety applications based on vehicle-to-vehicle (V2V) communication have become a major goal toward the intelligent transportation. The performance of the applications is mainly affected by the communication links, which in turn are governed by the topology characteristics of vehicular ad hoc networks (VANETs). To analyze the performance of V2V communication for VANETs, accurate spatial modeling is of great importance. There is an absence of a widely accepted two-dimensional model that well characterizes the random vehicle locations; the widely used Poisson point process (PPP) model is simple but not close to reality. In this paper, we concentrate on spatial point process modeling for random vehicle locations in large and small cities, performing empirical experiments with real location data of mobile taxi trajectories recorded by the global positioning system (GPS) in Beijing city of China and Porto city of Portugal. We find that the empirical probability mass functions (PMFs) of the number of taxis in test sets in different regions of Beijing or in Porto all follow a negative binomial (NB) distribution. The spatial correlations of the points are established by comparing the results of different sampling methods. Based on the above, we show that the Log Gaussian Cox Process (LGCP) model, whose empirical PMF nicely fits the NB distribution, accurately characterizes diverse spatial point patterns of random vehicle location in both large and small cities. This is verified by the minimum contrast method. Then we study the node degree as an important metric for the communication performance of the networks. It is shown that the connectivity of the LGCP model closely represents the connectivity found in the actual data, for both representative cities.**

The LGCP model is far more accurate than the widely used one-dimensional models and the 2-dimensional PPP for modeling the vehicle distribution, which is significant in V2V communication.

***Index Terms*—V2V, spatial point process, connectivity, VANET, modeling.**

I. Introduction

A. Motivation

Significant advances of wireless communications and the pervasive use of mobile electronics have turned vehicular networks from a futuristic promise to an attainable technology to meet the imminent demands for reduced accidents and improved road safety and efficiency. Safety applications in vehicular ad hoc networks (VANETs) will be based on the exchange of periodic cooperative awareness messages (CAMs). A CAM contains details such as a vehicle's location, speed, direction and other information that can be useful to drivers in the same region [1]. When an important event is detected by a vehicle (e.g., an accident, a slippery road), the information has to be rapidly delivered to other network nodes. Different from the conventional cellular communication, the network topology of vehicle-to-vehicle (V2V) communication changes greatly with the movement of vehicles. In turn, the connectivity of V2V also changes as a function of the topology (i.e., spatial vehicle distribution). To analyze the connectivity of VANETs, an accurate understanding of the topology and distribution is thus important. Stochastic geometry models offer a relevant view to location-dependent network characteristics. To describe the vehicular patterns and characterize the connectivity of vehicle-to-vehicle (V2V) communication, spatial stochastic models are crucial.

Q. Cui and N. Wang are with School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, 100876, China (email: cuiqimei@bupt.edu.cn).

Martin Haenggi is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA (email: mhaenggi@nd.edu).

B. Related Work

A road system is first added to wireless communication systems in [2] for strategic planning and economic analysis, and the Poisson line process is proposed to model the road position, and the Poisson point process (PPP) is used to model traffic on the road. In [3], the telecommunication networks including the road system are modeled as random geometric graphs, on the edges of which the locations of the network nodes are modeled as linear Poisson processes. Reference [4] observes that the vehicles that enter the highway through one of the traffic entry points form a Poisson process and studies the impact of vehicle mobility on the connectivity of VANETs. Connectivity is also analyzed in [5] and [6], where the same model for vehicles entering the highway is used and, in addition, it is assumed that the speed of the vehicles is constant or normally distributed. In [7] and [8], the vehicle locations are assumed to form a Poisson point process in one dimension, and the authors perform an analytical evaluation of broadcast protocols. In [9], the random locations of the vehicles in a street are modeled as a one-dimensional stationary Cox process with Fox's H-distributed random intensity. The inter-vehicle distances were simply set to certain values (i.e., from a few meters to 100 meters) to analyze delay in multiple V2V environment in [22]. [23] models a linear PPP for vehicles along a highway. It captures the influence of physical layer fading channels by deriving the joint distribution of the distance between every vehicle.

Besides the above studies with theoretic models, empirical research based on real data recently has focused on spatial point process modeling and analysis of real road systems or

random vehicle locations or traffic. For example, [10] fits Poisson line tessellations, Poisson-Voronoi tessellations, and Poisson-Delaunay tessellations to real data of the road system in Paris for the purpose of cost analysis and strategic planning of telecommunication networks. The authors in [11] further utilize random tessellations to model the road systems in fixed-access networks, and random points are added to the tessellation edges to represent network components. Reference [12] uses an inhomogeneous Poisson process model for vehicles at traffic signals and shows that the introduction of traffic signals does not affect the Poisson property of the stochastic model when the vehicles have a deterministic velocity profile, which is validated against empirical data in London city.

In [13], by analyzing real data of traffic counts in highway from the Bureau of Transport Statistics, New South Wales, the vehicles are modeled as a planar PPP, and numerical evaluation also indicates the proposed model is more accurate than existing one-dimensional models under sparse vehicle traffic densities. [14] verifies that the two-dimensional PPP assumption is only accurate for low vehicle density. In [15], experimental results for single-vehicle data of the Dutch freeway A9 and the German freeway A5 display that when the distances between cars are correlated due to traffic congestion, the vehicles do not follow a Poisson distribution.

The spatial point modeling for vehicles is of great importantance for infrustructure deployment and connectivity analysis in vehicular networks. Reference [31] presents a spatial model to study the load demand and allocates the charging stations based on the load demand. Reference [30] assumes a Poisson flow as a priori distribution to study the probabilistic distribution of inter-vehicle spacing and the connectivity in VANETs. Many works, e.g., [32]–[35], assume that the number of vehicles in a road section follows a Poisson distribution for simplicity. Reference [37] assumes that the vehicles arrive at the highway through one traffic entry point according to a Poisson process. It then studies the significance of the mobility on the connectivity of VANET by deriving the probability distribution of the node population size. Similarly, reference [38] assumes that the vehicles from one entry point follows a Poisson process, and determines the relation between the flow rate and the network connectivity. In reference [39], the connectivity is derived based on the inter-vehicle initial distance function. The connectivity of VANETs has attracted attention and interests of researchers in recent years [36], [40], [41]. The spatial distribution (especially the inter-vehicle distance) is crucial in the study of VANETs connectivity, which makes the spatial point process modeling for vehicles of great significance.

Based on the above, the previous research work has not produced a widely accepted spatial model that well characterizes the random vehicle locations, although the two-dimensional model is closer to reality. In urban VANET scenarios, it is of great importance to account for random vehicle locations in V2V communication. The locations of vehicles are correlated because of the traffic congestion and intersections. Most models ignore the fact that, in urban areas, there is strong correlation between the locations of vehicles. The widely used PPP model assumes that the points are completely independent, so it is not suitable for vehicle modeling, especially in big cities with heavy traffic. Therefore, spatial point process modeling of random vehicle locations based on real data analysis and mining is still a crucial problem to be addressed.

C. Contribution and Paper Organization

In this paper, we concentrate on spatial point process modeling and analysis of random vehicle locations in a large and a smaller city–Beijing city of China with a large-scale and regular road system, and Porto city of Portugal with a small and irregular road system. We perform empirical experiments with the global positioning system (GPS) traces of taxis in Beijing and Porto. The data of Beijing records realtime GPS location information for 12509 taxis over a month, and the data of Porto consists of realtime GPS location information for 442 taxis over a year. Our goal is to identify a planar point process to accurately model random vehicle locations using tools from stochastic geometry and to analyze the influence of spatial point model on the connectivity of V2V. The paper is an extension of the conference paper entitled "Spatial point process modeling of vehicles in large and small cities" [16].

Our main contributions are:

- The empirical probability mass functions (PMFs) of point counts of the point pattern generated by the vehicles are analyzed by different sampling methods. We find that the empirical PMFs of the number of taxis in test sets in different regions of Beijing or in Porto all follow a negative binomial (NB) distribution.
- The spatial correlations of the points are established by comparing the results of different sampling methods. The reduced Palm and reduced 2-Palm sampling methods shown in Fig. 4 and Table III are suitable to establish the clustering properties of the point pattern.
- We establish that the LGCP model, whose empirical PMF nicely fits the NB distribution, accurately characterizes diverse spatial point patterns of random vehicle locations where points are strongly clustered.
- Next we focus on the connectivity of V2V networks. The node degree is selected as a key metric to evaluate the V2V connectivity in this paper. The best spatial point process model is the model whose node degree distribution is the closest to that of the real vehicles. We verify that the LGCP model not only has a similar classic statistics with the real taxis, but also the closest connectivity performance. The LGCP model whose empirical PMF nicely fits the NB distribution is an accurate model for vehicle spatial distribution.

The rest of the paper is organized as follows. Section II presents the spatial models and characteristics used in the paper. Section III is devoted to the analysis of the empirical PMFs of the counting measure of the vehicle point process. We fit the Thomas cluster process, the Matérn cluster process, and the LGCP models to real vehicle location data by the minimum contrast method in Section IV. In Section V, we study the node degree and average node degree as additional criteria to decide on the best model for the spatial distribution of real vehicles. Finally, conclusions are drawn in Section VI.

II. Spatial Point Process Models and Characteristics

A. Spatial Point Process Models

There are several kinds of spatial point processes to describe a collection of random points in two dimensions, such as PPPs, cluster processes, hard-core processes, Cox processes, and Gibbs processes [17, Ch. 3]. Each of them has different characteristics—PPPs exhibit complete spatial randomness due to their independence property; Cox processes and other cluster processes are overdispersed relative to PPPs, i.e., they are more irregular; hard-core processes have a minimum distance between points and thus are more regular than the PPPs; Gibbs processes may be overdispersed or underdispersed. Vehicles are often clustered due to traffic congestion and intersections, so Thomas cluster processes, Matérn cluster processes, and LGCPs are promising candidates to model the vehicle pattern. The three candidate point process models—the Thomas cluster Process, the Matérn cluster process, and the LGCP—are described in this section.

A general cluster Poisson process is generated by taking a parent point process and daughter point processes for each parent, and translating the daughter processes to the position of their parent. The cluster process is then the union of all the daughter points. The Thomas cluster process and the Matérn cluster process are Neyman-Scott cluster processes, defined in [17, Ch. 3]. The Neyman-Scott process is a process whose parent process is a PPP and daughter points are random in number, independent of each other, and identically distributed.

1) Thomas Process: A doubly Poisson cluster process, where the intensity function of a cluster is given by

$$\lambda_0(x) = \frac{\bar{c}}{2\pi\delta^2} \exp\left(-\frac{\|x\|^2}{2\delta^2}\right), \tag{1}$$

i.e., the daughter points are normally scattered with variance δ^2 around each parent point, and the mean number of daughter point is $\bar{c}$.

2) Matérn Cluster Process: A doubly Poisson cluster process, where the intensity of the cluster can be expressed as

$$\lambda_0(x) = \frac{\bar{c}}{\pi R^2} \mathbf{1}_{b(o,R)}(x), \tag{2}$$

where $\mathbf{1}(.)$ is the indicator function. i.e., the daughter points are uniformly scattered on the ball of radius R centered at each parent point, and the mean number of daughter points is $\bar{c}$.

A general Cox process is a doubly stochastic Poisson process where the intensity measure itself is random. The intensity measure is a realization of a non-negative locally finite random measure [17, Ch. 3]. The Thomas and Matérn cluster processes are specific Cox processes, as is the LGCP, defined next.

3) Log Gaussian Cox Process (LGCP): A Cox process where the logarithm of the intensity function is a Gaussian process [20]. Specifically, the random intensity function of a LGCP is given as $\Lambda(s) = \exp\{Y(s)\}$ where $Y = \{Y(s) : s \in \mathbb{R}^2\}$ is a real-valued Gaussian process on the plane. Specifically, $Y(s) \sim \mathcal{N}(\mu, C)$, which means the random function Y is distributed as a Gaussian process with mean function μ and covariance function C. The exponential covariance can be parametrized in the form

$$C(r) = \beta \exp\left(-\frac{r}{\alpha}\right). \tag{3}$$

where r is the distance and β and α are parameters controlling the strength and the scale of autocorrelation respectively. The intensity of the LGCP is $\lambda = \exp(\mu + \beta/2)$.

B. Spatial Point Process Characteristics

For stationary point processes, there are five classical statistics called G, F, J, K and L functions to describe the inter-point dependence [17]. The G function is the nearest-neighbor distance distribution. The F function, also called the empty space function, is the cumulative distribution function of the distance from an arbitrary fixed location to the nearest point of the spatial point process. The J function is defined as $J(r) \triangleq \frac{1-G(r)}{1-F(r)}$, which is a measure of how close a process is to a PPP. Ripley's K function is defined such that $\lambda K(r)$ is the expected number of additional random points within a distance r of the typical point of the point process, where λ is the intensity of the process. The L-function is a transformation of Ripley's K function, defined as $L(r) \triangleq \sqrt{\frac{K(r)}{\pi}}$.

III. PMF Analysis of Real Location Data

A. Spatial Point Specifications

A single realization of a spatial point process is called a deterministic point pattern [17]. We treat the taxi locations as a realization of a point process and aim at finding the point process that has the highest likelihood of producing a realization containing the taxi locations of Beijing and Porto. The city of Beijing and the city of Porto are two representative cities with different traffic patterns and topologies.

1) Beijing: The data set of Beijing comprises GPS information of 12509 taxis for a month (from 2012/11/01 to 2012/11/27). It contains 785.4 million entries, each one comprised of the taxi location and metadata information associated with each taxi, as shown in Table I. The frequency of recording position information varies from one to six times per minute.

Beijing city is the capital of China with a population of 21 million on an area of 16,000 km^2; its road system is large, and relatively regular. The intensity of the taxis in Beijing is not constant and varies with the location due to the impact of the road systems. Fig. 1 shows the locations of 6927 vehicles in Beijing at 08:30 am on Nov. 2, 2012. From Fig. 1, a stationary point process appears unsuitable to accurately model the entire data set. Hence we partition the city into 9 regions in which the point pattern is relatively homogeneous. By doing so, we can fit a stationary point process model to each region. The points in Region 2 and Region 5 are densely distributed, the points in Region 9 exhibits strong clustering, and the patterns in the other regions are relatively sparse.

2) Porto: The data set of Porto includes GPS information of 442 taxis during a complete year (from 2013/07/01 to 2014/06/30). There are 1.7 million entries, of which each one represents a complete taxi trajectory, as shown in Table II.

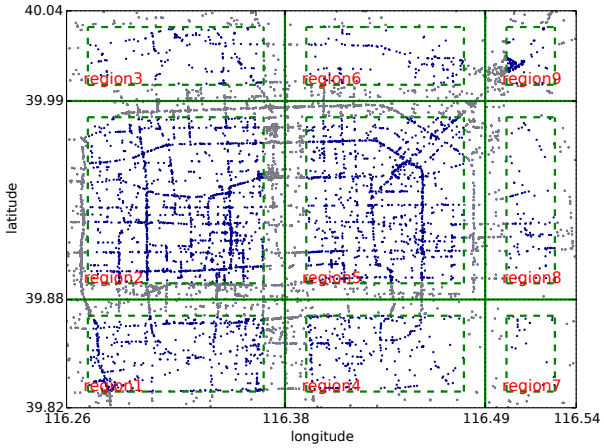

Fig. 1. The taxi distribution at 8:30 am on 2012/11/02 in 9 regions of Beijing, the horizontal axis represents the longitude from 116.26E to 116.54E and the vertical axis represents the latitude from 39.82N to 40.04N. The green dashed lines delineate the sampling window to which the test set centers are constrained.

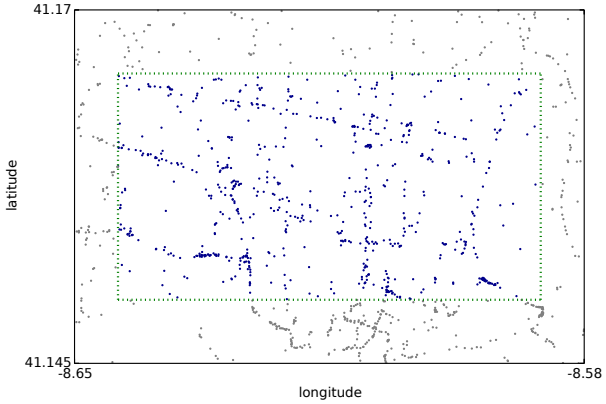

Fig. 2. The taxi distribution at 9:00 am (2013/07/01-2013/07/31) in Porto, the horizontal axis represents the longitude from 8.65W to 8.58W and the vertical axis represents the latitude from 41.145N to 41.17N. The green dashed lines delineate the sampling window to which the test set centers are constrained.

The complete taxi trajectory is a sequence of GPS positions measured every 15 seconds.

Porto is a small port city in Portugal with a population of 0.26 million on an area of 41 km^2; its road system is small and irregular. The spatial point pattern generated by taxi data of Porto is more homogeneous. A single realization is depicted in Fig. 2, which consists of 1245 taxi locations at 9:00 am over a month (2013/07/01 to 2013/07/31). Different from Beijing, we choose the union of 31 daily snapshots taken at 9:00 am because the number of points at 9:00 am on a single day is too small (less than 50) to be statistically significant.

TABLE I
FORMAT FOR BEIJING DATA SET

taxi ID	status	time	longitude	latitude	velocity
470341	0	20121102084536	116.5713	39.8063	39.8

TABLE II
FORMAT FOR PORTO DATA SET

trip ID	taxi ID	timestamp	polyline
T16	20000440	1408037740	[-8.618,41.136],[-8.618,41.135]$\cdots$

TABLE III
2-PALM SAMPLING METHOD

Algorithm: 2-Palm Sampling
Parameter Setting: a, b, d
// a, b: length and width of the rectangle, d: distance threshold
1. **while** $(x_1, y_1), (x_2, y_2)$ in φ **do**
2. $l \leftarrow \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$
3. **if** $l < d$
4. $N \leftarrow 0$ //initialize N
5. $s \leftarrow \frac{y_2 - y_1}{x_2 - x_1}$ //the slope
6. **for** (x, y) in φ
7. $b_1 \leftarrow \frac{(x_1 y_2 - x - 2y_1)}{x1 - x2} + \frac{bl}{2(x2 - x1)}$, Line1: $y = sx + b_1$
8. $b_1 \leftarrow \frac{(x_1 y_2 - x - 2y_1)}{x1 - x2} - \frac{bl}{2(x2 - x1)}$, Line2: $y = sx + b_2$
9. $b_3 \leftarrow \frac{al}{2} + \frac{x_1^2 - x_2^2 + y_1^2 - y_2^2}{2(y_1 - y_2)}$, Line3: $y = -x/s + b_3$
10. $b_4 \leftarrow -\frac{al}{2} + \frac{x_1^2 - x_2^2 + y_1^2 - y_2^2}{2(y_1 - y_2)}$, Line4: $y = -x/s + b_4$
11. **if** (x, y) in the rectangle generated by Line 1, 2, 3, 4
12. $N \leftarrow N + 1$
13. **end for**
14. **end while**

B. Empirical PMFs with Different Sampling Methods

In a spatial model, the countable random collection of the vehicles on the Euclidean space $\mathbb{R}^2$ is regarded as a point process $\Phi = \{x_1, x_2, \cdots\}$, where x_i denotes a vehicle location. The counting measure $\Phi(B) \triangleq \#\{\Phi \cap B\}$ is a random variable that denotes the number of vehicles in a Borel set $B \subset \mathbb{R}^2$ [17]. Similarly, we use the deterministic counterpart $\varphi(B) \triangleq \#\{\varphi \cap B\}$ to denote the counting measure of a point pattern φ.

To calculate the empirical PMFs of the point counts $\varphi(B)$, the first step is to sample and count the number of points generated by the taxi data. The basic principle of sampling and counting is to choose many test sets B_i so that they cover the entire region in a uniform manner, and count the number of points falling in B_i.

We consider four shapes—circle, square, rectangle, and cross rectangle—for sampling due to the irregularity of the real roads in cities. The cross rectangle is the union of a horizontally and a vertically oriented rectangle where the center square is excluded, as illustrated in Fig. 3. As shown in Table IV, the area of the test sets is fixed to 0.25 km^2 to maintain consistency. To eliminate boundary effects, we use a sampling window that is smaller than the region. The gap between the sampling window boundary and the region boundary is set as 1000 m for Beijing or 500 m for Porto shown in Fig. 1 and Fig. 2, so that the test sets do not exceed the region. The four sampling methods are:

1) Lattice Sampling: A sampling method in which the center of B follows a square-lattice distribution.

2) Uniform Random Sampling: A sampling method where the center of B follows a uniform PPP.

3) Reduced Palm Sampling: A sampling method in which B is centered at the location of the taxis.

4) Reduced 2-Palm Sampling: A sampling method where the center of B is at the midpoint of two taxi locations. The distance between any two taxis vary from 400 m to 800 m. The sampling method is shown in Fig. 4, and the algorithm is shown in Table III.

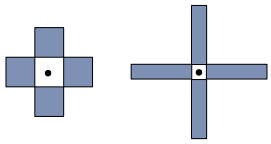

Fig. 3. The cross rectangle in the Palm sampling method. The left subfigure is the cross rectangle in which the ratio of arm length to arm width is 1, and the right subfigure is the cross rectangle in which the ratio of arm length to arm width is 4. Note that the center part is excluded.

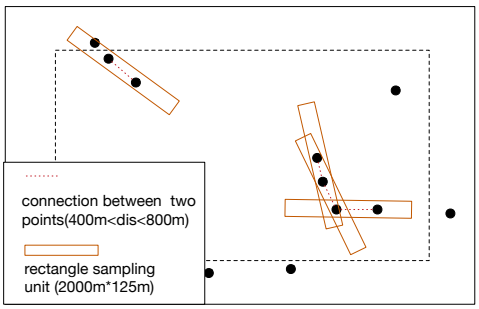

Fig. 4. The 2-Palm sampling methods. 'dis' in the legend refers to the distance between two points.

Fig. 5 shows the empirical PMFs and the corresponding cumulative distribution functions (CDFs) of the point counts with the lattice sampling and the uniform-random sampling methods in Region 1. Fig. 6 shows the results of the reduced Palm sampling methods based on four shapes and the reduced 2-Palm sampling method in Region 1. The Palm sampling and 2-Palm sampling are used to capture the impact of the underlying street system. Note that 'Palm-Xrectangle1' refers to Palm-sampling method for the cross-rectangle case in which the ratio of arm length to arm width is 1, while the ratio is 4 in 'Palm-Xrectangle4' as shown in Fig. 3. For Palm sampling, the cross rectangular shape of the test set is chosen to cover the street that the taxi lies on, given that most of the streets in Beijing are oriented north-south or east-west. For 2-Palm sampling, the rectangular shape is the optimal choice because thanks to the calculation of the orientation, the test set covers the street that the two taxis lie on.

The empirical PMF curves in Fig. 5 are monotonically decreasing, while the empirical PMF curves in Fig. 6 first increase and then decrease with increasing point number. This indicates that lattice sampling and uniform random sampling are definitely different from Palm and 2-Palm sampling and that different shapes have little impact on the empirical PMF calculation. Observing both results, we find that the empirical PMF of the point number less than 3 in Palm and 2-Palm sampling is smaller than that in lattice sampling and uniform random sampling because the former is constrained to sample in the vicinity of taxis. The same conclusion can be drawn for the other 8 regions. Moreover, random sampling and uniform random sampling are similar, and the shapes of the test sets barely affect the PMFs. Hence the following experiments only adopt the lattice sampling method.

The differences of the empirical PMF curves in Fig. 5 and those in Fig. 6 indicate that there are strong correlations among

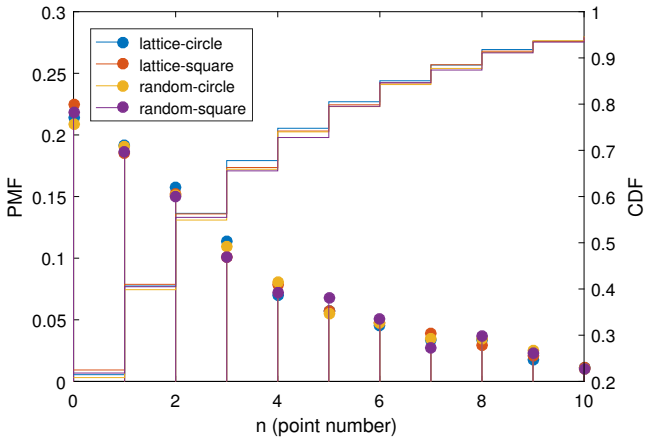

Fig. 5. The empirical PMFs and corresponding CDFs of point number in test set in Region 1 of Beijing with lattice and uniform random sampling methods.

vehicles, since the PMF curves of different sampling methods would coincide if the points were completely independent (i.e., from a PPP). The empirical PMF curves of a PPP are shown in Fig. 7. The ten PMF curves of the different sampling methods almost coincide. The existing difference is due to the random variability in generating a PPP realization and sampling it. The vehicles on roads can be sampled by the 2-Palm sampling method, without knowing the road network. The Palm and 2-Palm sampling methods are good choices to calculate the density of the points clustered on roads, since the boundaries of roads and non-road space are hard to obtain. The shape of the test set B in the Palm sampling method can be adjusted to the type of points to be sampled, and a rectangular test set is suitable for sampling vehicles on roads.

TABLE IV
SAMPLING METHODS

Sampling method	Shape of test set	Size of test set	Area of test set (km^2)
Lattice sampling	circle	radius=281 m	0.25
	square	500 m×500 m	0.25
Uniform random sampling	circle	radius=281 m	0.25
	square	500 m×500 m	0.25
Palm sampling	circle	radius=281 m	0.25
	square	500 m×500 m	0.25
	cross rectangle	250 m×250 m×4	0.25
		500 m×125 m×4	0.25
2-Palm sampling	rectangle	2000 m×125 m	0.25

C. Statistics on the Basis of the PMFs

On the basis of the empirical PMFs, we evaluate three statistics—empirical mean, variance, and intensity. We use φ to denote the point pattern generated by the taxi data. The empirical mean in Region i is given by

$$M = \sum_{k=1}^{K} \frac{\varphi_i(B_k)}{K}, \tag{4}$$

where φ_i is the data set of Region i, K is the number of test sets, and B_k denotes the kth test set.

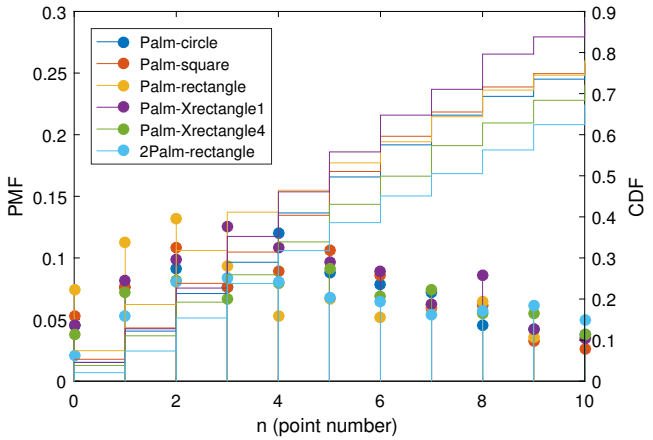

Fig. 6. The empirical PMFs and corresponding CDFs of point number in test set in Region 1 of Beijing with reduced Palm and reduced 2-Palm sampling methods.

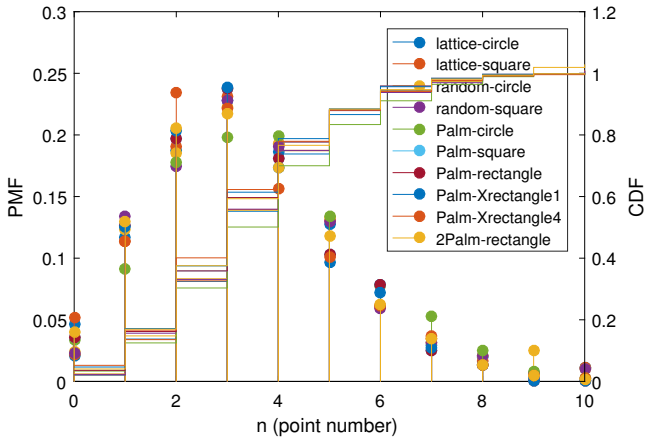

Fig. 7. The empirical PMFs and corresponding CDFs of point number in test set in a realization of a PPP with lattice, uniform random, reduced Palm and reduced 2-Palm sampling methods. The PPP is generated so that its intensity is the same as the mean density of vehicles in Region 1 of Beijing.

The variance is derived using the empirical mean as

$$\sigma^2 = \sum_{k=1}^{K} \frac{[\varphi_i(B_k)]^2}{K} - M^2. \tag{5}$$

The intensity is the expected number of points in a test set B, given by

$$\lambda = \frac{M}{|B|} = 4M/\text{km}^2. \tag{6}$$

TABLE V
MEAN, VARIANCE AND INTENSITY IN REGION 1 OF BEIJING

Sampling method	Shape of test set	Mean	Variance	Intensity (/km^2)
Lattice sampling	circle	3.22	13.48	12.89
	square	3.24	13.55	12.97
Uniform random sampling	circle	3.31	13.72	13.24
	square	3.32	13.99	13.26
Palm sampling	circle	6.93	28.3	27.73
	square	6.74	27.82	26.96
	cross	5.81	18.76	23.25
	rectangle	7.46	26.77	29.85
2-Palm sampling	rectangle	9.21	52.25	36.83

The results shown in Table V verify that the taxi points do not follow a Poisson distribution (whose mean and variance are the same). NB models are suitable for discrete data like count data when the variance is larger than the mean. The PMF of a NB random variable N is

$$\mathbb{P}(N=k) = \binom{k+r+1}{k} p^k (1-p)^r, \tag{7}$$

where $r > 0$ and $0 < p < 1$ are parameters of the NB distribution. Here, we attempt to fit the discrete empirical PMF data to the PMF of the NB distribution.

We use the maximum likelihood method to estimate the parameters r and p of the NB distribution. The experimental results show that the empirical PMFs in Regions 1 and 4 are very close, coinciding with those in Regions 2 and 5 and those in Regions 3, 6 and 7. Thus, only the empirical PMF data in Regions 1, 2, 3 and 8 with three different methods are provided in Fig. 8, Fig. 9, and Fig. 10 as typical representatives. Judging from the results, it is observed that the empirical PMFs in all the regions can fit NB distributions with specific r and p parameters well. The fitted NB distributions in Region 2 and Region 1 that belong to dense regions are both close under different sampling methods. Similarly, the fitted NB distributions in Region 3 and Region 8 that belong to sparse regions are approximate under lattice sampling method, but there is a significant discrepancy in the parameters for the Palm and 2-Palm sampling methods due to the distinctive sparsity in the different regions. Fig. 11 shows the empirical PMF data in Porto, which also fits the NB distribution very well.

***Remark 1**: The empirical PMFs of the number of points in test set in different regions of Beijing and in Porto follow a NB distribution, reflecting self-similarity in dense regions or in sparse regions and in large-scale and regular road systems or in small and irregular road systems.*

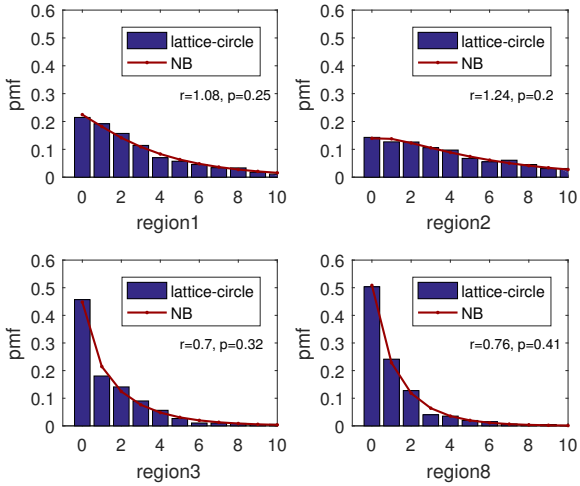

Fig. 8. Empirical PMFs of the point number in test set in Regions 1, 2, 3, 8 of Beijing with lattice-circle sampling, and fitted PMFs of NB variables.

IV. Fitting Method and Pattern Modeling

In this section, we develop a point process model whose statistics match those of the taxi data.

Based on the results we found in Section III-C, we fit the Thomas model, the Matérn cluster model, and the LGCP

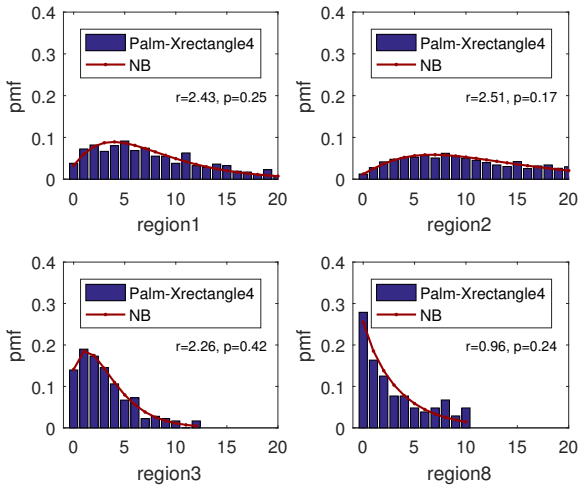

Fig. 9. Empirical PMFs of the point number in test set in Regions 1, 2, 3, 8 of Beijing with Palm-Xrectangle4 sampling, and fitted PMFs of NB variables.

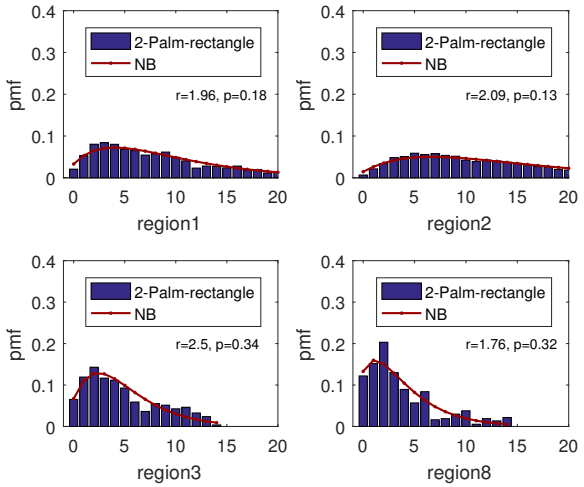

Fig. 10. Empirical PMFs of the point number in test set in Region 1, 2, 3, 8 of Beijing with 2-Palm-rectangle sampling, and fitted PMFs of NB variables.

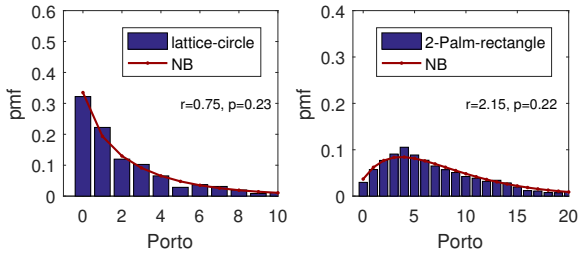

Fig. 11. Empirical PMFs of the point number in test set of Porto with lattice-circle and 2-Palm-rectangle sampling, and fitted PMFs of NB variables.

model to the given point set, using the minimum contrast method [18]. It fits the model by matching the data's summary statistic to its theoretical value. The K function describes the correlation between points which makes it a suitable summary statistic for fitting. In some cases, the K function of a point process is known exactly, as an analytic expression in terms of the model parameters. These cases include the Thomas and Matérn cluster processes, and the LGCP.

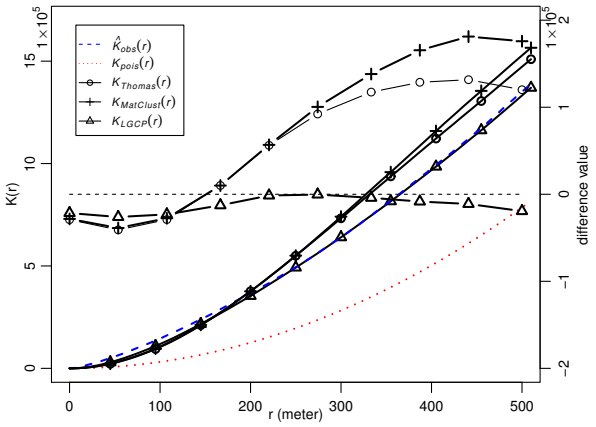

Fig. 12. Left axis: K function of the observed point pattern in Region 2 of Beijing and three fitted models. Right axis: the difference between the K functions of the observed point pattern and the other curves of three models.

For the ease of notation, the parameters of the point process models are collectively referred to as θ. For example, in the case of LGCP, θ refers to (α, μ, β). We determine the values of the parameters of a point process that give the closest match between the theoretical expected value of the summary statistic $K_\theta(r)$ and the observed value of the summary statistic evaluated from the data, denoted by $\hat{K}(r)$. The best match is determined by minimizing the discrepancy D between two functions, which is defined as

$$D(\theta) = \int_a^b \left| \hat{K}(r)^q - K_\theta(r)^q \right|^p dr, \tag{8}$$

where $a = 0$, $b = 500$ since the clustering of vehicles of which the distance are less than 500 meters are considered, and $p = 2$, $q = 1/4$ so that the contrast criterion is the integrated squared difference between the fourth roots of the K functions [19]. We employ a generic fitting algorithm for the method of minimum contrast, since the theoretical K function of the three models can be computed exactly from the model parameters.

Figs. 12, 13, and 14 compare the K function of the observed point patterns in Region 2 and Region 3 of Beijing, and Porto, to the K function of the fitted models, and depict the gap between the K function of the observed point pattern and the fitted models. The gap between the $K(r)$ curve of the observed point pattern and the curve of the LGCP model with the NB distributed PMF is the least, making the model a good fit for the data. The red dashed line denotes the K function of the fitted PPP model. The gap between the $K(r)$ curve of the observed point pattern and the curve of the PPP model is very large, showing that the PPP model is not suitable for the vehicle pattern. The same conclusion can also be drawn for the other regions of Beijing.

The LGCP model is uniquely determined by the parameters $\theta = (\alpha, \mu, \beta)$. The values of the parameters of the fitted LGCP model for Region 2 and Region 3 of Beijing and Porto are $(248.86, -11.65, 1.81)$, $(177.20, -13.30, 2.62)$, and $(105.56, 10.57, 2.22)$, respectively.

To model the locations of all vehicles, the density of the LGCP need to be adjusted. According to reference [29],

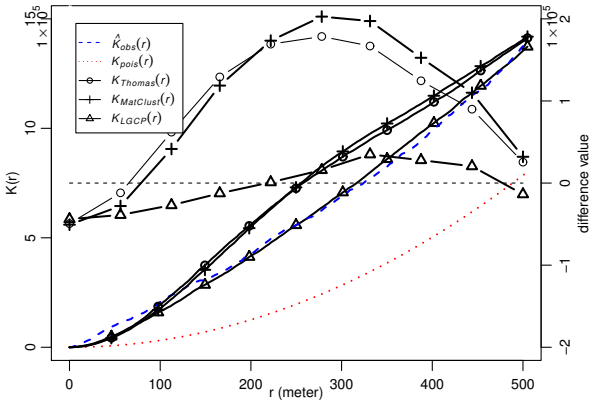

Fig. 13. Left axis: K function of the observed point pattern in Region 3 of Beijing and three fitted models. Right axis: the difference between the K functions of the observed point pattern and the other curves of three models.

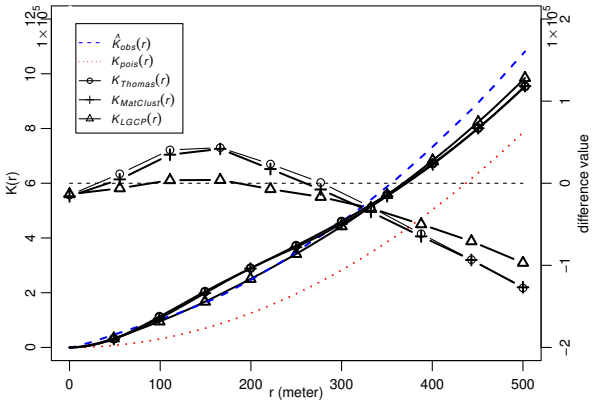

Fig. 14. Left axis: K function of the observed point pattern in Porto and three fitted models. Right axis: the difference between the K functions of the observed point pattern and the other curves of three models.

the density of all the traffic is about 20 times that of the data set used in our paper. In Section II-A, the density of the LGCP with parameters $\theta = (\alpha, \mu, \beta)$ is expressed by $\lambda = \exp(\mu + \beta/2)$. The parameter α is insensitive to the absolute density, since the spatial scale of the correlation in the underlying Gaussian field should be well captured already in a sparser data set.

We deduce that the LGCP model is an accurate model for the spatial point patterns of vehicle locations. Additionally, we perform a Monte Carlo test to determine the goodness-of-fit between the empirical and fitted models. In the Monte Carlo test, we regard the fitted LGCP model as the null hypothesis. To determine whether the vehicular point patterns follow the LGCP model, we compare the K functions of the data and the fitted LGCP model. We simulate realizations from the null hypothesis and compute the K function for each realization. The K function curves of the realizations form an envelope for a large number of simulations. The Monte Carlo test rejects the null hypothesis if the K function of the original spatial point pattern lies outside the envelope. As shown in Figs. 15, 16, and 17, the observed curve stays inside the envelope of the LGCP model fitted to the original point pattern, ascertaining that the original point pattern can be characterized by the

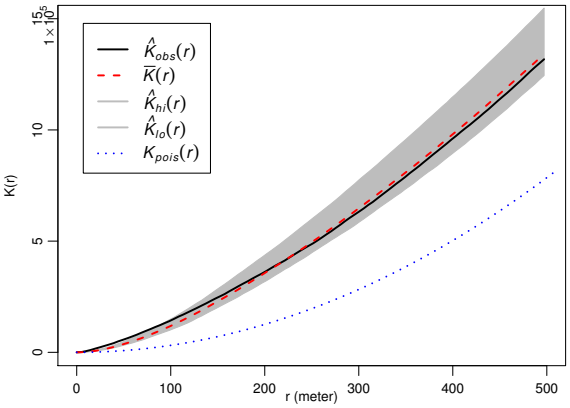

Fig. 15. Monte Carlo test on K function of taxis in Region 2 of Beijing, and the envelope of 39 realizations of the LGCP model. The red dashed line is the average value of the K function of 39 realizations of the fitted LGCP model.

LGCP model. Therefore, the LGCP model whose empirical point count PMF nicely fits a NB distribution can be used to model the vehicle locations. The K function of the vehicular point pattern in Region 2 of Beijing and the K function of vehicular point pattern in Region 3 of Beijing are similar, which further validates the self-similarity of the vehicular point pattern in the spatial domain.

Based on the above results, the LGCP model is more suitable for a spatial point distribution that is restricted to lines (i.e., the roads) than the Matérn and the Thomas cluster models. According to the definitions in (1) and (2), the Thomas and the Matérn cluster model are generated by placing daughter points around parent points. Nevertheless, the vehicles in cities are distributed along roads, not around some center points. The classical cluster point processes are not good choices for vehicle modeling, though the vehicles are overdispersed because of the intersection and congestions. The LGCP model is also suitable for the overdispersed vehicles, but without the above limitations of the Thomas and the Matérn cluster point processes.

Both the vehicle density and the road topology in Beijing and Porto are different, but the LGCP model is well fitted with the real vehicle distributions of the two cities. Based on the above, the LGCP model is valid also for other cities with different traffic patterns and topologies.

***Remark 2**: The LGCP model is more suitable for a spatial point distribution that is restricted to lines (for example, the roads) than the Matérn and the Thomas cluster models.*

V. Model Validation Based on VANET Connectivity

The LGCP model is verified as the best point process model for the vehicles in terms of the K function. It is sensible to use a statistic that is related with a standard metric used in wireless networks to decide on the best model [21]. In [21], the authors develop different spatial point process models for base stations in cellular networks based on empirical data, and propose coverage probability as the criterion for the goodness-of-fit. Here we study the node degree and the average node degree

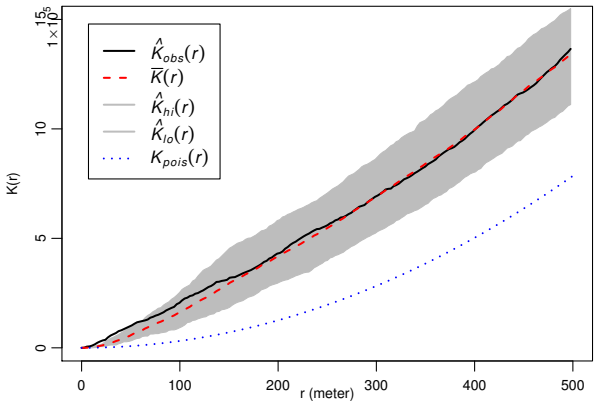

Fig. 16. Monte Carlo test on K function of taxis in Region 3 of Beijing, and the envelope of 39 realizations of the LGCP model. The red dashed line is the average value of the K function of 39 realizations of the fitted LGCP model.

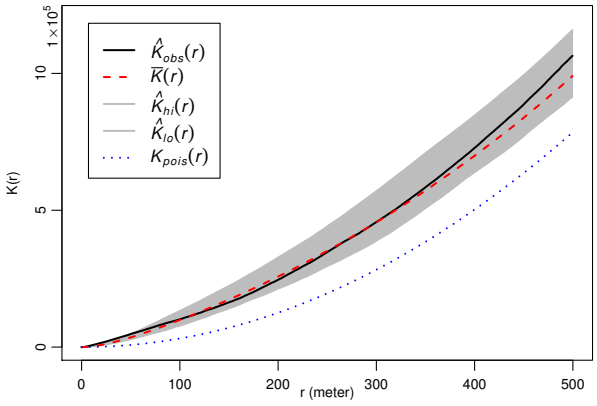

Fig. 17. Monte Carlo test on K function of taxis of Porto, and the envelope of 39 realizations of the LGCP model. The red dashed line is the average value of the K function of 39 realizations of the fitted LGCP model.

as additional metrics for the communication performance of the network to further validate the model.

The node degree is defined as the number of neighboring vehicles a vehicle can communicate with. We study the node degree distribution and the mean node degree. We compare the node degree of real taxis to simulated taxis with three different spatial point process models. It is noted that the parameters of the three cluster processes in this section are the same as those in Section IV. In other words, the LGCP, Matérn and Thomas parameters are not optimized based on the connectivity but based on the K function.

A. The Connectivity Model

In this section, a simple and general large-scale fading channel model, the log-normal model, is selected as the connectivity model. The log-normal model is used to account for the additional attenuation caused by obstacles in the V2V communications. These variations in this model is log-normal distributed as follows:

$$P_{\mathrm{rx}}(d) = P_0 - 10n\log_{10}\frac{d}{d_0} + N_\sigma, \tag{9}$$

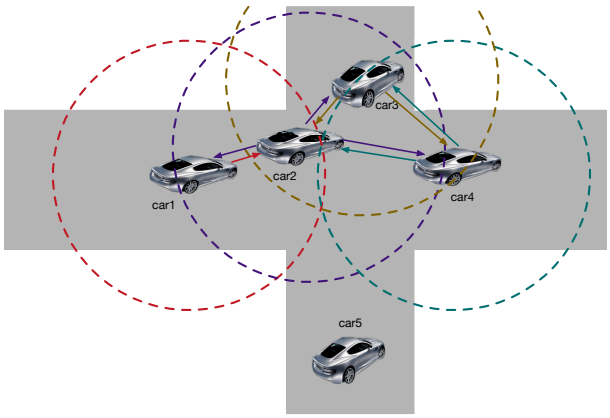

Fig. 18. Illustration of V2V communication network with one-hop wireless links.

where d is the distance between the transmitter and the receiver, d_0 denotes the reference distance, $P_{\mathrm{rx}}(d)$ is the received signal power at distance d (in dBm), P_0 is the received signal power at the reference distance d_0 (in dBm), n is the path loss exponent and N_σ is a zero mean Gaussian random variable with variance σ^2. A vehicle can communicate with another vehicle if $P_{\mathrm{rx}}(d)$ is greater than a certain threshold value P_{th}. Note that when $\sigma = 0$, the log-normal model reduces to the disk model. The disk model is a special case of the general log-normal model in which two vehicles are directly connected if and only if their Euclidean distance is less than or equal to r. Due to this fact, we set $P_{\mathrm{th}} = P_0 - 10n\log_{10}\frac{r}{d_0}$, so that the results in the log-normal model can be compared with the results in the disk model.

The connectivity function $g(d)$ is separately defined for the general case and disk model. It gives the probability that two vehicles at distance d are connected.

1) The Connectivity in the General Case: $g_{\mathrm{l}}(d)$ is used to denote the connectivity function of the log-normal model, which can be expressed as

$$\begin{aligned} g_{\mathrm{l}}(d) &= \mathbb{P}\left(P_{\mathrm{rx}}(d) > P_{\mathrm{th}}\right) \\ &= \mathbb{P}\left(P_0 - 10n\log_{10}\frac{d}{d_0} + N_\sigma > P_0 - 10n\log_{10}\frac{r}{d_0}\right) \\ &= \mathbb{P}\left(N_\sigma - 10n\log_{10}\frac{d}{r} > 0\right) \\ &= Q\left(\frac{10n}{\sigma}\log_{10}\frac{d}{r}\right), \end{aligned} \tag{10}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{y^2}{2}\right)dy$ is the tail probability of the standard normal distribution.

2) The Connectivity in the Disk Model: The disk model, a special case of the log-normal model, has the connectivity function $g_{\mathrm{d}}(d) = \mathbf{1}(d \le r)$, where r is predetermined, commonly known as the transmission range. This model is widely used in the analysis of the VANET topology characteristics due to its simplicity [22]–[26]. We adjust the radius r to the overall density of vehicles in the region, such that $\lambda\pi r^2$ is roughly the same for different regions. This allows a better comparison of the CDFs across the different regions. For example, since the vehicle densities in Region 2, Region 3 and Porto are

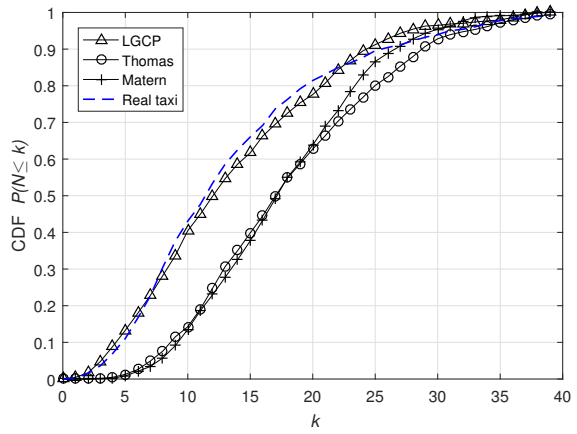

Fig. 19. The cumulative density function of the node degree in Region 2 of Beijing for the log-normal model ($\sigma = 5$ dB, $n = 2.5$).

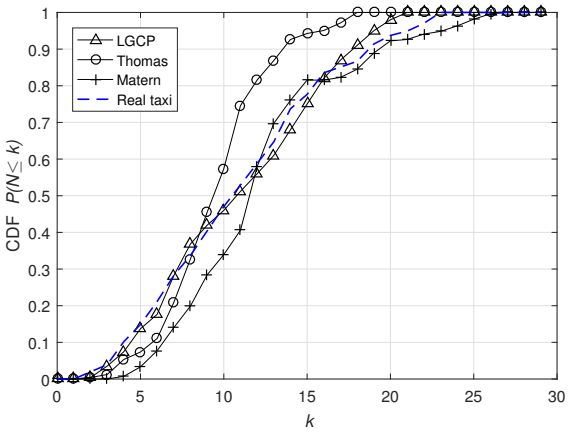

Fig. 20. The cumulative density function of the node degree in Region 3 of Beijing for the log-normal model ($\sigma = 5$ dB, $n = 2.5$).

19.32/km^2, 5.99/km^2, and 78.35/km^2, respectively, the radii r are set with 300 m, 538 m, and 149 m in these regions. These radii are consistent with IEEE 802.11p, which is designed for V2V and vehicle-to-infrastructure communications, allows transmission ranges up to 1,000 m [27].

The communication model is a broadcast model where vehicles communication with each other with one-hop wireless links as illustrated in Fig. 18. A vehicle sends a message in broadcast mode, and vehicles can receive the message if the distance between transmitter and receiver are less than the distance threshold for the disk model. Vehicles can receive the message if the received power is larger than the power threshold for the log-normal model.

B. Node Degree Distribution with Fixed Transmission Range

When the transmission range of the vehicles is fixed, the node degree will be only influenced by the relative position of vehicles. Because of the vehicles' inhomogeneous distribution, the node degree of different vehicles is different. We focus on the statistical characteristic of the node degree (i.e, the cumulative distribution function of the node degree) of vehicles at a certain time.

1) Node Degree in the General Case: In the log-normal model, vehicle x can communicate with vehicle y if and only if the received power $P_{\rm rx}\left(\|x-y\|\right)$ is greater than the threshold value $P_{\rm th}$, where $\|x-y\|$ denotes the distance between vehicle x and vehicle y. The node degree of vehicle x in the log-normal model is defined as the node degree averaged over the shadowing, and it can be expressed as

$$N(x,r) = \sum_{y\in\varphi} Q\left(\frac{10n}{\sigma}\log_{10}\frac{\|x-y\|}{r}\right), \qquad (11)$$

where φ denotes the countable set of taxis, n is the path loss exponent, σ is the standard variance of the Gaussian random variable, and r is the mean transmission range.

The parameter values of n and σ of the model are chosen based on the channel measurement results reported in [24], [25], and [26]: $n = 2.5$, $\sigma = 5$ dB. And the CDFs of the node degree in Region 2 and Region 3 of Beijing with the log-normal model are illustrated in Fig. 19 and Fig. 20, and the mean transmission range is equal to the transmission range in the disk model to have a fair comparison. We compare the node degree of real taxis to simulated taxis with three different spatial point process models. It turns out that for the LGCP model, the discrepancy between the node degree of real taxis and the node degree of the spatial model is the smallest. The same conclusion can be drawn in the other regions of Beijing city and Porto city (shown in Fig. 21). Accordingly, we conclude that the LGCP model is the best model when considering the connectivity of the VANET.

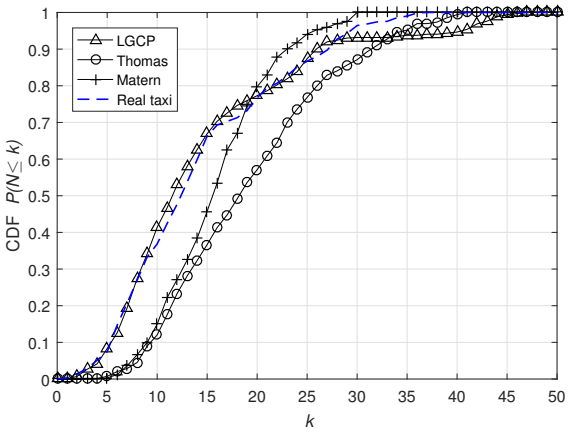

Fig. 21. The cumulative density function of the node degree in Porto for the log-normal model ($\sigma = 5$ dB, $n = 2.5$).

2) Node Degree in the Disk Model: The node degree of vehicle x in the disk model with transmission range r is

$$N(x,r) = \#\{\varphi \cap b(x,r)\} - 1, \qquad (12)$$

where φ denotes the countable set of taxis, $\#\{.\}$ denotes the number of elements in a set, and $b(x,r)$ denotes the disk of radius r centered at x. We can use the reduced Palm sampling method in Section III-B to determine the empirical cumulative density function (CDF) of the node degree $N(x,r)$.

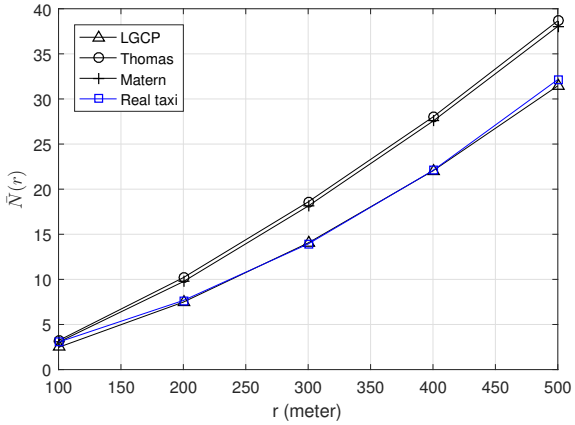

Fig. 22. The average node degree in Region 2 of Beijing for the log-normal model.

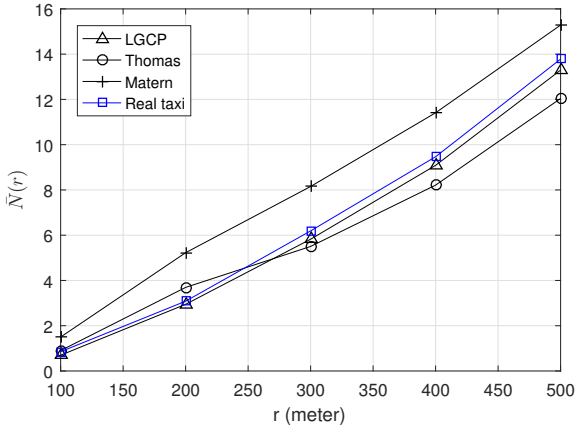

Fig. 23. The average node degree in Region 3 of Beijing for the log-normal model.

C. Average Node Degree

When the transmission range of the vehicles increases, the node degree of each vehicle will increase or stay unchanged. The average node degree is defined as the mean of the node degree of all the vehicles.

1) Average Node Degree in the General Case: The average node degree in Region i in the log-normal model can be expressed as

$$\bar{N}_i(r) = \frac{1}{\#\{\varphi_i\}} \sum_{x \in \varphi_i} N(x, r), \tag{13}$$

where $\#\{.\}$ denotes the number of elements in set, φ_i denotes the countable set of taxis in Region i, $N(x, r)$ is the node degree of vehicle x in the log-normal model with power threshold $P_{\text{th}} = P_0 - 10n \log_{10} \frac{r}{d_0}$, $N(x, r)$ can be calculated using (11).

The average node degree in Region 2 and Region 3 of Beijing city in the log-normal model are illustrated in Fig. 22 and Fig. 23. We compare the average node degrees of real taxis to simulated taxis with three different spatial point process models. It shows that the difference between the average node degree of real taxis and average node degree of simulated taxis with the LGCP model is the smallest. Fig. 22 also shows that the average node degree of the real taxis is smaller than the average node degree of the Thomas and Matérn models. It is verified that the connectivity of the LGCP model is closest to the connectivity of vehicles in both small (shown in Fig. 24) and large cities.

2) Average Node Degree in the Disk Model: For a disk model (i.e., a special case of the log-normal model), the average node degree in Region i in the disk model with transmission range r can be expressed as

$$\bar{N}_i(r) = \lambda K(r), \tag{14}$$

The average node degree for the Thomas cluster point process Φ_{t} with parameters $\bar{c}, \kappa, \delta$ can be expressed as [28]

$$\bar{N}_{\text{t}}(r) = \bar{c}\kappa \left(\pi r^2 + \frac{1}{\kappa} \left(1 - \exp\left(-\frac{r^2}{4\delta^2} \right) \right) \right), \tag{15}$$

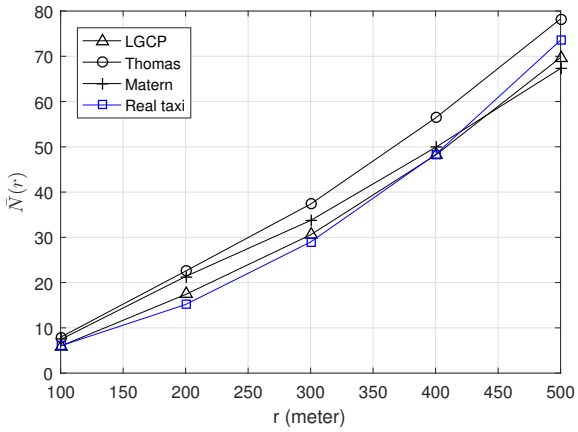

Fig. 24. The average node degree in Porto for the log-normal model.

where κ is the intensity of the parent point process.

The average node degree for the Matérn cluster point process Φ_{m} with parameters $\bar{c}, \kappa, R$ can be expressed as [28]

$$\bar{N}_{\text{m}}(r) = \bar{c}\kappa \left(\pi r^2 + \frac{1}{\kappa} h\left(\frac{r}{2R} \right) \right), \tag{16}$$

where κ is the intensity of the parent point process, R is the radius of the disk inside which the offspring points are distributed around the parent, and

$$h(z) = 2 + \frac{1}{\pi}[(8z^2 - 4)\arccos z - 2\arcsin z + 4z\sqrt{(1 - z^2)^3} - 6z\sqrt{1 - z^2}] \tag{17}$$

for $z \leq 1$, and $h(z) = 1$ for $z > 1$.

The average node degree of the LGCP Φ_{l} with parameters μ, α, β can be expressed as [28]

$$\bar{N}_{\text{l}}(r) = \exp\left(\mu + \frac{\beta}{2} \right) \int_0^r 2\pi s \exp\left(\sigma^2 e^{\frac{-s}{\alpha}} \right) ds. \tag{18}$$

The average node degree and the K function of the pattern of vehicles are proportional to each other, as shown in (14). Since the empirical K functions of the vehicular point patterns in Beijing city and Porto city have been illustrated in Figs. 12, 13, and 14, the average node degrees of the vehicles

in the disk model are not reproduced.

VI. Conclusion

We adopted various sampling methods to study the distribution of vehicles in Beijing, with large-scale and regular road systems, and Porto, with small-scale and irregular road system. We showed that the number of vehicles in both the cities follows a NB distribution through an analysis of the probability mass functions. The spatial correlations of the points are detected by comparing the results of different sampling methods. The vehicles on roads can be conditionally sampled by the 2-Palm sampling method, without knowing the road network. The reduced Palm and reduced 2-Palm sampling methods are good choices for sampling points constrained to lines.

To find a suitable point process for the vehicular network, we fitted the Thomas process, Matérn cluster process, and the LGCP with the PMF following a certain NB distribution to the given point set using the minimum contrast method. It is shown that the LGCP model whose empirical PMF nicely fits the NB distribution can generally characterize the spatial point pattern of random vehicle locations very well in both large and small cities. We use a Monte Carlo test on the classical statistics as the criterion for goodness-of-fit, which proves that the LGCP model performs better in fitting than other models. The LGCP model is more suitable for a spatial point distribution that is restricted by lines (for an example, the roads), than the Matérn cluster model and Thomas cluster model. Additionally, we propose the node degree and average node degree as additional criteria to decide on the best model for the spatial distribution of real vehicles in V2V communication scenario. The results verify that the LGCP model can accurately characterize the vehicle spatial distribution in V2V communication.

Accordingly, we can use the LGCP model to analyze other performance metrics (i.e., coverage and capacity) and optimize the practical deployment of VANETs. Equipped with an accurate model for the vehicle positions, the powerful tools of stochastic geometry [17] can be applied to analyze the communication performance of V2V and V2I networks. This entails choosing a communication model (which vehicles communicate) and a model for path loss and fading and the determining the distribution of the signal-to-interference-and-noise ratio (SINR) at the receiver.

Acknowledgment

The work was supported by National Nature Science Foundation of China Project (Grant No. 61471058, 61461136002) and the 111 Project of China (B16006). The partial support of the U.S. National Science Foundation through grant CCF 1525904 is gratefully acknowledged. We also thank Jeya Pradha for her constructive comments on earlier drafts of the manuscript.

References

[1] R. Stanica, E. Chaput and A. L. Beylot, “Enhancements of IEEE 802.11p Protocol for Access Control on a VANET Control Channel," in *2011 IEEE International Conference on Communications (ICC)*, Kyoto, 2011, pp. 1-5.

[2] F. Baccelli, M. Klein, M. Lebourges, et al., “Stochastic geometry and architecture of communication networks," *Telecommun. Syst.*, vol. 7, pp. 209-227, Jun. 1997.

[3] F. Voss, C. Gloaguen, F. Fleischer, et al., “Distributional properties of Euclidean distances in wireless networks involving road systems," *IEEE J. Sel. Areas in Commun.*, vol. 27, no. 7, pp. 1047-1055, Sep. 2009.

[4] M. Khabazian and M. K. M. Ali, “A performance modeling of connectivity in vehicular ad hoc networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2440-2450, Jul. 2008.

[5] V. K. Muhammed Ajeer, P. C. Neelakantan, and A. V. Babu, “Network connectivity of one-dimensional vehicular ad hoc network," in *2011 Int. Conf. Commun. Signal Processing*, Calicut, India, 2011, pp. 241-245.

[6] S. Yousefi, E. Altman, R. El-Azouzi, et al., “Analytical model for connectivity in Vehicular Ad Hoc Networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3341-3356, Nov. 2008.

[7] S. Busanelli, G. Ferrari, and R. Gruppini, “Performance analysis of broadcast protocols in VANETs with Poisson vehicle distribution," in *11th Int. Conf. ITS Telecommun.*, St. Petersburg, Russia, 2011, pp. 133-138.

[8] A. Busson, “Analysis and simulation of a message dissemination algorithm for VANET," *Int. J. Commun. Syst.*, vol. 24, no. 9, pp. 1212-1229, Jan. 2011.

[9] Y. Jeong, J. W. Chong, H. Shin, et al., “Intervehicle communication: Cox-Fox modeling," *IEEE J. Sel. Areas in Commun.*, vol. 31, no. 9, pp. 418-433, Sep. 2013.

[10] C. Gloaguen, F. Fleischer, H. Schmidt, et al., “Fitting of stochastic telecommunication network models, via distance measures and Monte-Carlo tests," *Telecommun. Syst.*, vol. 31, is. 4, pp. 353-377, Apr. 2006.

[11] C. Gloaguen, F. Voss, and V. Schmidt, “Parametric distance distributions for fixed access network analysis and planning," in *Proc. 21st Int. Teletraffic Congress*, Paris, France, 2009, pp. 1-8.

[12] I. W. H. Ho, K. K. Leung, and J. W. Polak, “Stochastic model and connectivity dynamics for VANETs in signalized road systems," *IEEE/ACM Trans. Networking*, vol. 19, no. 1, pp. 195-208, Feb. 2011.

[13] P. Golmohammadi, P. Mokhtarian, F. Safaei, et al., “An analytical model of network connectivity in vehicular ad hoc networks using spatial point processes," in *Proc. of IEEE Int. Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*, Sydney, New South Wales, Australia, 2014, pp. 1-6.

[14] F. L. Mannering, S. S. Washburn, and W. P. Kilareski. *Principles of highway engineering and traffic analysis*, 4th Edition. Wiley, 2008.

[15] A. Y. Abul Magd, “Modeling highway-traffic headway distributions using superstatistics," *Phys. Rev. E*, vol. 76, pp. 057101.1-057101.4, Nov. 2007.

[16] Q. Cui, N. Wang, and M. Haenggi, “Spatial point process modeling of vehicles in large and small cities,” *IEEE Global Communications Conference*, Singapore, 2017.

[17] M. Haenggi, *Stochastic geometry for wireless networks*. Cambridge University Press, 2012.

[18] P. J. Diggle, and R. J. Gratton, “Monte Carlo methods of inference for implicit statistical models,” *J. Royal Statist. Soc.*, vol. series B 46, no. 2, pp. 193-212, Jan. 1984.

[19] R. P. Waagepetersen, “An estimating function approach to inference for inhomogeneous Neyman-Scott processes,” *Biometrics*, vol. 63, pp. 252-258, Apr. 2007.

[20] J. Møller, A. R. Syversveen, and R. P. Waagepetersen, “Log Gaussian Cox processes”. *Scand. J. Statist.*, vol. 25, no. 3, pp. 451-482, Sep. 1998.

[21] A. Guo and M. Haenggi, “Spatial stochastic models and metrics for the structure of base stations in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5800-5812, Nov. 2013.

[22] D. W. Matolak, I. Sen and W. Xiong, “Channel modeling for V2V communications," in *2006 3rd Annual Int. Conf. on Mobile and Ubiquitous Systems-Workshops*, San Jose, CA, 2006, pp. 1-7.

[23] H. Wang, R. P. Liu, W. Ni, W. Chen, and I. B. Collings, “VANET modeling and clustering design under practical traffic, channel and mobility conditions," *IEEE Trans. on Commun.*, vol. 63, no. 3, pp. 870-881, Mar. 2015.

[24] L. Cheng, B. E. Henty, F. Bai, and D. D. Stancil, “Highway and rural propagation channel modeling for vehicle-to-vehicle communications at 5.9 GHz,” in *IEEE Antennas and Propagation Society Int. Symposium*, July 2008, pp. 1-4.

[25] L. Cheng, B. E. Henty, D. D. Stancil, F. Bai, and P. Mudalige, “Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band,” *IEEE J. Sel. Areas Commun.*, vol. 25, no. 8, pp. 1501-1516, Oct. 2007.

[26] O. Renaudin, V.-M. Kolmonen, P. Vainikainen, and C. Oestges, “Non-stationary narrowband MIMO inter-vehicle channel characterization in the 5 GHz band,” *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 2007-2015, May 2010.

[27] M. Slavik and I. Mahgoub, “Spatial Distribution and Channel Quality Adaptive Protocol for Multihop Wireless Broadcast Routing in VANET," *IEEE Transactions on Mobile Computing*, vol. 12, no. 4, pp. 722-734, Apr. 2013.

[28] J. Møller and R. P. Waagepetersen, *Statistical inference and simulation for spatial point processes*, Chapman & Hall/CRC, Boca Raton, 2003.

[29] H. Liu, C. He, J. Lents, N. Davis, M. Osses, and N. Nikkila, “Beijing Vehicle Activity Study,” tech. rep., International Sustainable Systems Research Center (ISSRC), Jan. 2005. Available at http://www.issrc.org/ive/downloads/reports/BeijingChina.pdf

[30] JJ. Guo, Y. Zhang, X. Chen, S. Yousefi, C. Guo and Y. Wang, “Spatial Stochastic Vehicle Traffic Modeling for VANETs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 2, pp. 416-425, Feb. 2018.

[31] I. Morro-Mello, A. Padilha-Feltrin and J. D. Melo, “Spatial-Temporal model to estimate the load curves of charging stations for electric vehicles," in *2017 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, Quito, 2017, pp. 1-6.

[32] C. Shao, S. Leng, Y. Zhang, A. Vinel, and M. Jonsson, “Performance analysis of connectivity probability and connectivity-aware MAC protocol design for platoon-based VANETs,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 12, pp. 5596-5609, Dec. 2015.

[33] S. Panichpapiboon and W. Pattara-Atikom, “Connectivity requirements for self-organizing traffic information systems,” *IEEE Trans. Veh. Technol.*, vol. 57, no. 6, pp. 3333-3340, Nov. 2008.

[34] W. Zhang et al., “Multi-hop connectivity probability in infrastructure-based vehicular networks,” *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 740-747, May 2012.

[35] M. Rios, V. Marianov, and M. Pérez, “Locating fixed roadside units in a bus transport network for maximum communications probability,” *Transp. Res. C, Emerg. Technol.*, vol. 53, pp. 35-47, Apr. 2015.

[36] S.-I. Sou and O. K. Tonguz, “Enhancing VANET connectivity through roadside units on highways,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3586-3602, Oct. 2011.

[37] M. Khabazian and M. K. M. Ali, “A performance modeling of connectivity in vehicular ad hoc networks,” *IEEE Trans. Veh. Technol.*, vol. 57, no. 4, pp. 2440-2450, Jul. 2008.

[38] S. Panichpapiboon and W. Pattara-Atikom, “Connectivity requirements for vehicular networks with single-hop broadcasting,” in *Proc. IEEE 9th Int. Conf. Intell. Transp. Syst. Telecommun. (ITST)*, Oct. 2009, pp. 388-392.

[39] R. Chen, Z. Zhong, and M. Ni, “Performance analysis on conditional connectivity for vehicular ad hoc networks,” in *Proc. 23rd IEEE Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sydney, NSW, Australia, Sep. 2012, pp. 641-668.

[40] H. Zhao, S. Chen, D. Li, and H. Zhu, “The establishment of the network connectivity model in VANET," in *2016 8th International Conference on Wireless Communications & Signal Processing (WCSP)*, Yangzhou, China, 2016, pp. 1-6.

[41] L. Zhang, L. Cai, J. Pan, and F. Tong, “A New Approach to the Directed Connectivity in Two-Dimensional Lattice Networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 11, pp. 2458-2472, Nov. 2014.