Concrete Synthetic Modeling of Vehicular Networks as Random Geometric Graphs
(Submission to ACM PE-WASUN 2011)

ABSTRACT
Random graphs are often used to model vehicular networks. However, their applicability has been limited because it is difficult to express and instantiate parameters of graph models of vehicular networks using real-life data. In this paper, we consider using random geometric graphs to model vehicular networks where vehicle movements are constrained to a road system. We show that vehicles form a random geometric graph with edge probability $p$ that can be expressed as a closed-form expression or as an algorithmically computable expression with parameters that are known or easily measurable in real life. This enables one to answer essential questions, such as questions related to routing and placement of roadside radio transceivers required to support vehicular communications, as a function of practically measurable and computable parameters.

1. INTRODUCTION
Vehicular communication networks are recently emerging as an area attracting an impressive amount of research. This fact is witnessed by the increasing number of related conferences and journals (see, e.g., [1]) and research projects (see, e.g., [2]). Reasons behind this fact include importance of applications (e.g., traffic safety, traffic awareness, ‘drive-through’ commercial transactions) as well as understanding that the technology needed to build vehicular networks seems reasonably mature. Realizing vehicular networks will undoubtedly be a hard task. It is expected to require combined research, development and policy efforts from multiple disciplines, including communication networks, mathematical modeling, autonomous and distributed computing, human factors, security and privacy, etc. In this paper we use concrete and synthetic mathematical modeling to study general frameworks for answering practical questions related to various aspects of vehicular networks, including communication protocols, routing and security.

Since the theory of random graphs was introduced by Erdos. P and Renyi. A, a large number of results have been achieved [3, 4, 5] by mathematical modeling of real-life networks as random graphs for more tractable investigations. Geometric random graphs are one of the most important variants of random graphs, with applications in various areas, including medicine, ecology, statistical physics and hypothesis testing (see, e.g., [6]), and ad-hoc, sensor and vehicular networks (see, e.g., [7, 8, 9, 10, 11]).

Vehicular networks have been modeled as a geometric random graph by representing vehicles as vertices and the communication links between two vehicles as edges. The graph changes with time and a snapshot of the graph at any time is a result of a) movement of vehicles following a mobility model, b) vehicles being restricted to a fixed geography and c) a communication model defining connections between nodes. Previous works that have used random graphs to model vehicular networks have either started by assuming that vehicles form a random graph or provided little justification. To the best of the authors’ knowledge, there is no paper in which the application of random geographic graphs to model vehicular networks is mathematically justified.

Our contribution. In this paper we start by defining a geographical model such as a Manhattan Grid and deploying vehicles which follow a mobility model such as the Manhattan Grid Mobility model. We then show that the graph formed by the vehicles at any point of time, $t$ is a random geometric graph with $n$ nodes and edge probability $p$. We show how to calculate the parameter $p$ via a closed form under relatively restricted assumptions first. Then, we generalize the analysis to algorithmically calculate $p$ for any communication model and geographical map. Several mobility models can be used to model vehicular networks [12, 13]. We extend our analysis to show which class of mobility models imply a graph which can be represented as a random geometric graph with sufficient accuracy. From a technical standpoint, the main contributions in this paper are: (1) proving that, given specific communication, geographical, and mobility models, the vehicles form a random geometric graph with and edge probability $p$ that can be expressed in closed form; (2) proving that, given general communication, geographical, and mobility models, the vehicles form a random geometric graph with and edge probability $p$ that can be algorithmically computed. From a conceptual standpoint, the main contributions in this paper are: (1) defining general variants of communication, geographical, and mobility models that will still result in a random geometric graph; (2) discussing a methodology that simplifies the analysis of
problems over vehicular networks, given their obtained random geometric graph characterization.

We believe our results provide partial justification towards the use of random geometric graphs as models of vehicular networks. Our synthetic modeling of vehicular networks as random geometric graphs satisfies a special concreteness property: the graph parameters can be computed as a function of the network’s real-life parameter values. Thus, further studies modeling vehicular networks as random geometric graphs can analyse protocols or applications on the random geometric graph abstraction of their networks and then translating their analysis results to their network using our closed-form or algorithmically computable expressions.

As an example of this methodology, we investigated applications of our results to two important problems in vehicular networks (which partially motivated this paper’s research): (1) survivable routing: specifically, how to ensure that messages are successfully routed in the presence of faulty nodes or nodes captured by attackers; (2) node placement to detect malicious vehicles: specifically, how many mobile surveillance nodes will be needed so that each node has a target number of neighbours that can help detecting misbehaving vehicles. We show that the average value of $p$ gives a concretely computable estimate of the metrics of interest for these two problems: the survivability of the vehicular network and the degree of confidence in malicious vehicle detection via neighbour voting. We stress that both survivable routing and node placement to detect malicious vehicles are of critical importance to ensure the security of all conventional vehicular network applications, such as traffic safety, traffic awareness and commercial transactions.

**Organization of the paper.** The rest of the paper is organized as follows: In Section 2 we present formal definitions of specific geographic, mobility and communication models and prove that a random geometric graph with closed-form computable parameters is obtained. Section 3 generalizes this result by generalizing the communication, geographical and mobility models and obtaining algorithmically computable parameters of the random geometric graph. Section 4 describes example applications to routing and security, and sections 5 discusses conclusions and future work.

## 2. CLOSED-FORM GRAPHS

In this section we present our first main technical result. We formulate geographic, mobility and communication models, and show that these induce the existence, at any given time, of a random geometric graph among the vehicles, where the edge parameter can be expressed using a closed-form formula. In what follows, we start by defining the geographic, mobility and communication models; we then analyse the spatial distributions of vehicles over time, and finally show that these distributions induce a random geometric graph.

### 2.1 Specific Models

Our model formulation, and the various modeling and simplification choices, are based on synthetic modeling approaches [13] that use mathematical modeling and do not rely on simulators, surveys or traffic traces. We add a dimension of concreteness in that our results can be expressed in closed-form formulas with practically obtainable parameter values.

**Geographic model.** The geographic model defines the set of possible vehicle positions.

We consider a $m \times m$ perfect grid; in other words, a square, divided into $m^2$ subsquares of equal area. The grid’s $m + 1$ horizontal lines, also called rows, and $m + 1$ vertical lines, also called columns, represent 2-way roads. Each side of a sub-square is divided into $s$ units. The geographic model is assumed to be static.

Let $N$ denote the number of total vehicles (or nodes) and let $n = n(t)$ denote the number of active (i.e., currently moving) vehicles at any given time $t$. For $i = 1, \ldots, N$, the $i$-th node is associated with a position, represented, at time $t$ as $P_i(t) = (x_i(t), y_i(t))$, where either $x_i(t) \in \{0, s, 2s, \ldots , ms\}$ and $y_i(t) \in \{0, \ldots , m\}$ (on horizontal lines), or $x_i(t) \in \{0, \ldots , m\}$ and $y_i(t) \in \{0, s, 2s, \ldots , ms\}$ (on vertical lines).

The initial positions of all nodes, denoted as $P_i(0)$, for $i = 1, \ldots, N$, are chosen randomly and independently among the above feasible values.

**Mobility model.** The mobility model defines how the vehicle positions evolve over time. It models driving patterns, multiple trips made by the same vehicle, and the length, starting point, and driving direction of each trip.

**Driving pattern:** We assume that each vehicle repeatedly goes through a drive phase followed by a pause phase. Each drive phase consists of a trip. For $i = 1, \ldots, N$, the $i$-th vehicle is assumed to make $dt$ trips per day.

**Multiple trip correlation:** We assume that the length, starting point and direction of any two trips from the same vehicle are independent random variables.

**Single trip length, starting point and direction:** For $i = 1, \ldots, n$, and $j = 1, \ldots, dt$, the $j$-th daily trip by the $i$-th active vehicle is assumed to start at a uniformly chosen point on the grid, and to follow a direction specified by (one instance of) the Manhattan Grid Mobility Model (MGMM) [14]. The latter can be seen as a modified discrete random walk on the grid determined by the grid defined in the geographic model, the modification consisting of a particular set of constraints on choosing the direction.

For $i = 1, \ldots, n$, the $i$-th active vehicle’s direction at time $t$ is $\text{dir}(t) \in \{U, D, L, R\}$ (for up, down, left, right), where $\text{dir}(t) \in \{L, R\}$ on horizontal lines and $\text{dir}(t) \in \{U, D\}$ on vertical lines. We consider the wrap-around model [15], where a node moving outside of the grid on one of the four sides enters the grid on the opposite side. The initial directions of all vehicles, denoted as $\text{dir}(0)$, for $i = 1, \ldots, n$, are chosen randomly and independently among all feasible values. We assume a discrete time scale, where at each time step the vehicles perform one movement step along their directions, according to the MGMM mobility model.

Specifically, at an intersection, the vehicle can turn left, right, go straight or take a u-turn with a certain probability. The vehicle is not allowed to change its direction when it is on a line segment connecting two intersection points. This restriction captures the fact that in real life vehicles...
are either not allowed or much less likely to take u-turns in between traffic signal (or cross-over) points. The formal definition of this mobility model is obtained by defining the evolution of \( P_i(t), dir_i(t) \), for all \( i = 1, \ldots, n \), and all \( t \geq 1 \) (the case \( t = 0 \) being defined above), as follows:

1. if \( (x_i(t) \equiv 0 \mod{s}) \land (y_i(t) \equiv 0 \mod{s}) \), then \( dir_i(t+1) \) is uniformly chosen among \( \{U, L, D, R\} \) else \( dir_i(t+1) = dir_i(t) \);
2. if \( (x_i(t) \equiv 0 \mod{s}) \) and \( dir_i(t) = U \), then:
   \( P_i(t+1) = (x_i(t), y_i(t)+1) \);
3. if \( (x_i(t) \equiv 0 \mod{s}) \) and \( dir_i(t) = D \), then:
   \( P_i(t+1) = (x_i(t), y_i(t)-1) \);
4. if \( (y_i(t) \equiv 0 \mod{s}) \) and \( dir_i(t) = L \), then:
   \( P_i(t+1) = (x_i(t)-1, y_i(t)) \);
5. if \( (y_i(t) \equiv 0 \mod{s}) \) and \( dir_i(t) = R \), then:
   \( P_i(t+1) = (x_i(t)+1, y_i(t)) \).

**Communication model.** The communication model defines the conditions under which any pair of vehicles can or cannot communicate.

As most commonly done in the literature, in this section we define a vehicle’s communication range or coverage area as a circle of radius \( r \), with the node as its center. When two vehicles are within each other’s communication range, we say there is an edge between them. Let \( V \) be the set of all vehicles in drive phase and let \( E(t) \) be the edges between these vehicles at time \( t \). Then, the communication graph \( G(t) = (V,E(t)) \) at time \( t \) is defined as the graph formed by nodes in \( V \) and edges such that \( (i,j) \in E(t) \) if and only if \( \| (P_i(t), P_j(t)) \| \leq r \), where \( \cdot \) is the Euclidean distance metric.

We also define the **neighbour set** of a node \( i \) in his drive phase at time \( t \), as the set of nodes in \( V_i \), which are in \( i \)'s communication range at time \( t \); more formally defined as \( N_i(t) = \{ j \in V : \| (P_i(t), P_j(t)) \| \leq r \} \).

### 2.2 Spatial Distribution of nodes

The spatial distribution of nodes defines the distribution, at any given time \( t \), of the position of each vehicle within the geographic model. The spatial distribution is determined by the vehicles’ initial positions at time \( 0 \) and their movements through the \( t \) steps based on the mobility model.

In this section we show that given the previously defined geographic, mobility and communication model, there exists a spatial distribution of the nodes that is stationary (i.e., it does not vary with the time \( t \)) regardless of the changes due to the mobility model and regardless of the distribution of the vehicles’ initial positions.

More formally, given a probability space, we define a (discrete-time) stochastic process \( S \) as a collection of random variables defined over the probability space and indexed by a discrete time variable; i.e., \( S = \{S(t) \mid t = 0, 1, 2, \ldots\} \). We say that a discrete-time stochastic process is stationary if for all integers \( k \geq 0 \), all integers \( \tau \geq 0 \), the joint distribution of \( (S(1+\tau), \ldots, S(k+\tau)) \) does not depend on \( \tau \). Now, consider the 2-dimension random variable \((X, Y)\) distributed according to the following distribution with parameter \( P > 0 \):

\[
f_{XY}(x, y) = \begin{cases} 
1/P & \text{for } (x, y) \in H_1 \\
2/P & \text{for } (x, y) \in H_2 \\
0 & \text{otherwise, where} 
\end{cases}
\]

Here \( H_2 \) can be seen as the set of ‘intersection’ nodes (e.g.: points \( t_2 \) and \( t_3 \) in Figure 1), \( H_1 \) as the set of ‘internal’ nodes (e.g.: point \( t_1 \) in Figure 1), and \( H \) as the set of all nodes on the grid. Given \( s, m \), and by direct counting over the grid, closed-form expressions can be derived for \( |H_1|, |H_2| \) and \( |H| \), as follows:

\[
|H_1| = 2(s-1)m(m+1), |H_2| = (m+1)^2, \quad \text{and} \quad |H| = |H_1| + |H_2| = 2sm^2 + 2sm - m^2 + 1.
\]

Then, a closed-form expression can be found for \( P \) by observing that \((1/P)|H_1| + (2/P)|H_2| = 1\), and thus obtaining \( P = (|H_1| + 2|H_2|) = 2(sm + 1)(m+1) \).

Now, let \( S_i \) denote the discrete-time stochastic process defined as \( S_i(t) = (X_i(t), Y_i(t)) \) for all integers \( t \geq 0 \). Here, \( S_i(t) \) represents the position of a node \( i \) at time \( t \). Our result on the spatial distribution of nodes is as follows.

**Theorem 1.** In the probability space given by the mobility model of Subsection 2.1, let \((S_1, \ldots, S_n)\) denote the discrete-time stochastic process describing the positions of all \( n \) nodes. If \( S_i(0) = f_{X_i,Y_i}(x, y) \) for \( i = 1, \ldots, n \), then \((S_1, \ldots, S_n)\) is stationary.

**Proof.** We note that the MMGM mobility model acts independently on each node. Then, by the definition of stationary stochastic processes, if each process \( S_i \) is stationary then so is process \((S_1, \ldots, S_n)\). The proof that process \( S_i \) is stationary is done by induction over \( k \), where the base case \( k = 1 \) is proved by induction over \( t \).

Let \( k = 1 \). We need to prove that \( S_i(t) = f_{X_i,Y_i}(x, y) \) for all \( t \geq 0 \). We prove this fact by induction over \( t \). The base case \( t = 0 \) directly follows by the theorem’s hypothesis. Now, assume that this fact holds up to \( t = u \); that is, \( S_i(t) = f_{X_i,Y_i}(x, y) \) for \( t = 1, \ldots, u \), and consider \( S_i(u+1) \). We have that \( S_i(u+1) = a \) for \( a \in H_1 \) either with
probability \((1/P)\times(1/2) + (1/P)\times(1/2) = 1/P\) (in correspondence of points like point \(O\) in Figure 1) or with probability \((2/P)\times(1/4) + (1/P)\times(1/2) = (1/P)\) (in correspondence of points like points \(A,B,C,D\) in Figure 1), and \(S(n + 1) = a\) for \(a \in H_2\) with probability \(4\times(1/P)\times(1/2) = 2/P\) (in correspondence of points like point \(O\) in Figure 1). This proves the claim when \(k = 1\).

Now, assume that the claim holds for \(k \leq q\), thus implying that the distribution of \((S_1(1 + \tau), \ldots, S_k(q + \tau))\) does not depend on \(\tau\). We want to show that the distribution of \((S(1 + \tau), \ldots, S(q + 1 + \tau), S(q + 1 + \tau))\) does not depend on \(\tau\). Because of the induction hypothesis, this will not happen only if the distribution of \((S_1(q + 1 + \tau), \ldots, S_k(q + \tau), S(q + 1 + \tau))\), conditioned by \(S_1(1 + \tau), \ldots, S_k(q + \tau)\), is independent on \(\tau\). However, note that, by definition of \(f_{XY}\), \(S_i(q + 1 + \tau)\) only depends on \(S_i(q + \tau)\) and thus the distribution of \((S_1(q + 1 + \tau), \ldots, S_k(q + \tau), S(q + 1 + \tau))\), conditioned by \(S_1(1 + \tau), \ldots, S_k(q + \tau)\) can be written as the distribution of \((S_i(q + 1 + \tau), \ldots, S_k(q + \tau))\). This latter distribution is independent on \(\tau\), or otherwise the induction hypothesis for \(q = 2\), is contradicted. □

Remark. Theorem 1 could have been proved using results from the theory of Markov chain. We preferred the above proof for sake of simplicity and self-containment.

2.3 Number of grid points covered by a circle

In this subsection we consider the set of grid points, denoted as \(C(x, y)\), covered by a circle of radius \(r\) and center at position \((x, y)\), and calculate the following three numbers: the total number of grid points covered by the circle, \(|C(x, y)|\), the number \(|C(x, y) \cap H_1|\) of internal points and the number \(|C(x, y) \cap H_2|\) of intersection points, in the circle with center \((x, y)\). These 3 numbers are critical quantities in the later calculation of the random geometric graph parameter \(p\).

In Fig. 2 we show a circle representing the communication range of a vehicle, whose center is displaced from its nearest intersection by \(d\) units horizontally. The value of \(d\) can be calculated as, \(d = \lfloor \min((x \mod s), s - (x \mod s)) + \min((y \mod s), s - (y \mod s)) \rfloor\). In vehicular networks, the vehicles are restricted to move on the roads and as a result one of the two terms on the right-hand side of the latter equation is always zero. Then the entire analysis shown below also holds for a vertical displacement, due to symmetry. We will also use the following notations:

- \(q_i, \beta = \sqrt{r^2 - ((i \times s)^2)}\),
- \(q_i, \alpha = \sqrt{r^2 - ((i - 1)s + d)^2}\), and
- \(q_i, \gamma = \sqrt{r^2 - ((i \times s) - d)^2}\).

The main part of the procedure to calculate \(|C(x, y)|\) consists of determining the number of points on each chord of the circle. The chords are differentiated as horizontal chords and vertical chords. The points covered by the horizontal chords remain the same with any amount of displacement on that axis. To calculate the number of points on horizontal chords we first calculate the value of \(c\). We observe that \(c = q_i, \alpha\) for \(i = 1\). The total horizontal chord length covered by the circle is \((2c + 1)\). There will be \(2\lfloor r/s \rfloor\) such horizontal chords and one diameter which is \(2r + 1\) in length. To calculate the number of points on vertical chords we first assume that there are \(\alpha\) vertical lines on the left and \(\beta\) vertical lines on the right of (or possibly including) the center of the circle. The length of the first vertical chord on the left is \((2a + 1)\), where \(a = q_i, \alpha\) for \(i = 1\). The length of the first vertical chord on the right is \((2b + 1)\), where \(b = q_i, \alpha\) for \(i = 1\). The sum of all the points on the horizontal lines and all vertical lines will include the intersection points twice. Then, subtracting the number of intersection points in the circle from the total sum gives the total number of points covered by the circle, \(|C(x, y)|\), which can be written as:

\[
|C(x, y)| = 4 \sum_{i=1}^{\lfloor \frac{r}{s} \rfloor} q_i, \alpha + 2 \lfloor \frac{r}{s} \rfloor + 2r + 1 + 2 \sum_{i=1}^{\alpha} q_i, \alpha + \alpha + 2 \sum_{i=1}^{\beta} q_i, \beta - |k_2|,
\]

where values \(\alpha, \beta\) will be later calculated. The number of intersections that the first vertical line on the right (resp., left) of the circle's center can accommodate is \(2 \lfloor b/s \rfloor + 1\) (resp., \(\lfloor a/s \rfloor + 1\)). Then the total number of intersection points in the circle, \(|C(x, y) \cap H_2|\), is equal to the sum of all the intersections on all the vertical chords, which is

\[
2 \sum_{i=1}^{\alpha} q_i, \alpha - \alpha + 2 \sum_{i=1}^{\beta} q_i, \beta + \beta.
\]
As we use a wrap-around model, and thus the communication range can be thought as always within the boundaries of the grid, the number of internal points in the circle is:

\[ |C(x, y) \cap H_1| = |C(x, y)| - |C(x, y) \cap H_2|. \]

The values of \( \alpha \) and \( \beta \) depends on the amount of displacement, \( d \), and the relationship between \( r \) and the distance between intersection points, \( s \), and is detailed in the next subsection, while computing a closed-form expression for \( p \).

### 2.4 Edge Probability

The value \( p \) can be computed as the sum, over all points \((x, y)\) in the grid, of the probability, denoted as \( \eta(x,y) \), that a first node is on point \((x,y)\) of the grid, times the probability, denoted as \( \xi(\alpha, \beta,d) \), that a second node is in the circle with radius \( r \) and center \((x,y)\). In what follows, we use the results in Section 2.3 and Section 2.2 to compute probabilities \( \xi(\alpha, \beta,d) \), \( \eta(x,y) \), and to sum over all grid points \((x,y)\).

Using the spatial distribution of nodes from Section 2.2 and this distribution’s stationarity, as from Theorem 1, we have

\[
\xi(\alpha, \beta,d) = \left( \frac{1}{P} \cdot |C(x, y) \cap H_1| + \frac{2}{P} \cdot |C(x, y) \cap H_2| \right).
\]

By substituting the values of \( |C(x, y) \cap H_1| \) and \( |C(x, y) \cap H_2| \) from Section 2.3, we obtain:

\[
P \cdot \xi(\alpha, \beta,d) = \left( \frac{4}{s} \sum_{i=1}^{\lfloor \frac{s}{r} \rfloor} q_{i,0} + 2 \left[ \frac{r}{s} \right] + 2r + 1 \right) + 2 \sum_{i=1}^{\alpha} q_{i,-} + \alpha + 2 \sum_{i=1}^{\beta} q_{i,+} + \beta - |k_2|.
\]

Now, consider the case where \( s \) is odd (the other case being similar and discussed later) and the following three subcases:

**Subcase 1:** If \((r \ mod \ s) = \lfloor s/2 \rfloor \) then, for any \( d \), we have that \( \alpha_1 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \) and \( \beta_1 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \). The summation needed to compute \( p \) is split into two addends, the first one over the internal points, and the second one over the intersection points. We finally obtain that

\[
p = \frac{1}{P} \cdot \frac{|H_1|}{s - 1} \left( 2 \cdot \sum_{d=1}^{\lfloor \frac{s}{r} \rfloor} \xi(\alpha_1, \beta_1,d) \right) + \frac{2}{P} \cdot |H_2| \cdot \xi(\alpha_1, \beta_1,0)
\]

\[
= \frac{4m(m+1)}{P} \sum_{d=1}^{\lfloor \frac{s}{r} \rfloor} \xi(\alpha_1, \beta_1,d) + \frac{2(m+1)^2 \xi(\alpha_1, \beta_1,0)}{P}.
\]

**Subcase 2:** If \((r \ mod \ s) < \lfloor s/2 \rfloor \) then,

- for \( d > \psi \), we have \( \alpha_2 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \) and \( \beta_2 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \),

where, \( \psi = (r \ mod \ s) + 1 \), and, similarly as for subcase 1, we obtain that

\[
p = \frac{4m(m+1)}{P} \left( \sum_{\psi=1}^{\lfloor \frac{s}{r} \rfloor} \xi(\alpha_1, \beta_1,\psi) + \xi(\alpha_2, \beta_2,\psi) \right) + \frac{2(m+1)^2 \xi(\alpha_1, \beta_1,0)}{P}.
\]

**Subcase 3:** If \((r \ mod \ s) > \lfloor s/2 \rfloor \) then,

- for \( d \leq \psi \), we have \( \alpha_1 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \) and \( \beta_1 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \);

- for \( d > \psi \), it holds that \( \alpha_2 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \) and \( \beta_2 = \left[ \frac{1}{2} \left( \frac{2r}{s} \right) \right] \),

where, \( \psi = (s - (r \ mod \ s) - 1) \), and \( p \) is computed as in subcase 2.

When \( s \) is even, subcase 1 does not exist and the condition of subcase 3 is modified as ‘If \((r \ mod \ s) \geq \lfloor s/2 \rfloor \)’.

### 2.5 Random geometric graphs: result

In this subsection, we formally state our result that, at any time \( t \), the communication graph \( G(t) = (V,E(t)) \) is a random geometric graph; i.e., it includes \( n \) nodes, any two nodes are connected if and only if their positions are at distance less than \( r \), and any two nodes are connected with the same probability \( p \). The crucial important addition provided by our result is that the parameters \( n,p \) can be expressed as a closed formula of parameters \( m, s, r \). Formally, we have:

**Theorem 2:** Assuming the initial placement of the \( n \) nodes follows the distribution \( f_{XY} \), and given the grid geographic model, the mobility model, and the communication model of Subsection 2.1 at any given discrete time \( t \geq 0 \), the communication graph \( G(t) = (V,E(t)) \) is a random geometric graph with \( n \) nodes and edge probability \( p \), where \( p \) has the closed-form expression of Subsection 2.4.

Because of the stationary distributions of the nodes guaranteed by Theorem 1, proving Theorem 2 only reduces to calculating \( p \), which is done in Subsection 2.4, respectively. Using computer calculations, we observed that the expression for \( p \) can be reasonably well approximated as

\[
p \approx \frac{1}{sm^2} \left( \sum_{i=1}^{\lfloor \frac{s}{r} \rfloor} \sqrt{r^2 - (i \times s)^2} + r \right).
\]

Specifically, in our calculation (based on several concrete values for parameters \( s, r, n, m \) and on random choices for the vehicles’ initial positions) we observed that this simpler formula was almost always an upper bound for the actual expression for \( p \), the error being typically smaller by about 1 or 2 orders of magnitude.

From the above equation, we can observe that when \( r \approx s \), \( p \) is \( \approx \frac{cs^2}{s^2m^2} \) which means \( p \) is proportional to the ratio of
area of the communication circle to the area of the total grid and when \( r < < s \), \( p \approx \frac{\pi r^2}{s^2} \), which means \( p \) is proportional to the length of the 2 circle chords from the grid.

3. GRAPHS UNDER GENERAL MODELS

In this section we present our second main technical result. In the previous section we have shown that given specific geographical, mobility and communication models, the communication graph is a random geometric graph where the edge parameter can be expressed as a closed-form expression in terms of known or measurable parameters from the models. In this section we extend this result by generalizing each one of the models considered in the previous section, and obtain that the communication graph is a random geometric graph where the edge parameter can be algorithmically computed (with efficient running time).

3.1 Generalizing Models

Generalizing the communication model. Given a fixed geographical and mobility model, such as those in Section 2.1, we now consider generalizing the communication model and studying the consequences on the communication graph. As often done in the literature, the preferred modeling of the communication range is by a circular shape. It is however interesting to see that the analysis assuming a circular coverage area generalizes to any arbitrary coverage area. Indeed, in the case of a squared coverage area, we were still able to obtain a closed-form expression (as a function of \( m, r, s \)) for the graph edge parameter \( p \) (omitted here). More generally, one could consider a communication range as an arbitrary 2-dimension shape. Since the spatial distribution of nodes on the grid only depends on the mobility and geographical model, such a change in the communication range of the nodes does not affect the computation of value \( p \) (as a function of \( m, s \)) or the stationarity of the spatial distribution of nodes (as calculated in Theorem 1). Because the stationarity of this distribution suffices to obtain a random geometric graph with \( n \) nodes and the same edge probability \( p \) in Theorem 2, we would still obtain a random geometric graph even when the communication range is an arbitrary 2-dimension shape. What changes here is the value of \( p \), which depends on the values of \( k_1, k_2 \), which directly depend on the communication range shape. However, a procedure properly generalizing to an arbitrary 2-dimensional shape the counting-based procedure used in Section 2.3 for the circle, would suffice to compute the new \( p \) value.

Generalizing the geographic model. Given a fixed mobility model, such as the one in Section 2.1, and a fixed communication model, such as the above generalized one, we now consider generalizing the geographical model and studying the consequences on the communication graph.

While the MMGM mobility model has often been considered in the research literature, in practice one would like to have a geographical model of an arbitrary street map. Indeed any such map can be easily abstracted as a planar graph \( G_{map} = (V_{map}, E_{map}) \), where \( V_{map} \) is the set of all intersections on the street map, and \( E_{map} \) is the set of all streets joining any two intersections. Moreover, to each edge in \( E_{map} \) one could associate a weight proportional to the street length (thus further generalizing our constant parameter \( s \) in the grid). Starting from \( G_{map} \), we can define a related graph \( G'_{map} = (V_{map}, E_{map}) \), as follows. The set \( V_{map} \) contains \( V_{map} \) as well as, for each \( (a, b) \in E_{map} \) with weight \( w \), nodes \( u_1, \ldots, u_{w-1} \). The set \( E_{map} \) contains, for each \( (a, b) \in E_{map} \) with weight \( w \), edges \((a, u_1), (u_1, u_2), \ldots, (u_{w-1}, b)\).

The approach used in Section 2.1 to find a stationary distribution over the grid generalizes to an arbitrary graph \( G'_{map} \), where the distribution is as follows:

\[
\sum_{(x, y) \in V_{map}} f_{XY}(x, y) = 1,
\]

for some value \( P \) such that

\[
f_{XY}(x, y) = \begin{cases} 
\frac{\deg(v_{x,y})}{P} & \text{for } (x, y) \in V'_{map} \\
0 & \text{otherwise},
\end{cases}
\]

and where \( v_{x,y} \) is a node from \( V'_{map} \) having position \((x, y)\) and \( \deg(v_{x,y}) \) is its degree in \( G_{map} \). This means that the probability of finding a node at a point \( v_{x,y} \) is proportional to the number of directions that point can be reached at.

Now, with the new stationary distribution, similar analysis as in Section 2.5 can be used to calculate the new edge probability \( p \). Specifically, we have that

\[
p = \sum_{v_{i,y}} pr(i, v_{x,y}) \cdot pr(i, j, v_{x,y}),
\]

where \( pr(i, v_{x,y}) \) is the probability that node \( i \) is on point \( v_{x,y} \), \( pr(i, j, v_{x,y}) \) is the probability that node \( j \) is connected to node \( i \), given that the latter is on point \( v_{x,y} \), and

\[
pr(i, j, v_{x,y}) = \sum_{x,y \in Z_i} f_{XY}(x, y).
\]

Thus, nodes moving with Manhattan mobility and fixed communication model will still form a random geometric graph on the generalized geographical model defined above, although with different values for the edge parameter \( p \).

Generalizing the mobility model. Given fixed communication and geographic models, such as the above generalized ones, we now consider generalizing the mobility model and studying the consequences on the communication graph. Our goal is to generalize the mobility model and still guarantee the existence of a stationary spatial distribution of nodes, as the latter will imply a random geometric graph, as in the proof of Theorem 2. We now show that the MGMM mobility model can be significantly generalized, and the generalization will allow the existence of a unique stationary distribution such that, regardless of the initial node deployment, the spatial distribution converges to this distribution.

In general terms, one could define a mobility model as an arbitrary probabilistic function that at any given time, given the entire history of the nodes’ movements on the geographic model, returns the next nodes’ movements. This definition turns out to be too general for our goal, and we restrict it...
so that the movements of each node only depends on a finite number of positions of the same node. Given the above generalized geographic and mobility model, a finite state Markov chain can be constructed as follows. First, assume the (simpler) mobility model only saying that the movement of each node only depends on the current position of the same node. Then the Markov chain’s states represent the points on the geographical model and transition probabilities are directly defined by the mobility model. Since any point on a map can be reached from any other point on the map, the Markov chain formed by mapping points on the map to states hold the same property. This makes the Markov chain irreducible. The extension to the more general mobility model where the movements of each node depends on a finite number of positions of the same node is obtained by a standard technique in the area of Markov chains that blows up the original state space $V$ into a (much larger) space $\mathbb{V}^n$. The necessary and sufficient condition for such a Markov chain to have a stationary distribution is obtained from the following known fact about Markov chains (see, e.g. [16]).

**FACT 1.** Assume an irreducible Markov chain is time-homogeneous, in that its transition matrix is time-independent. Then this chain has a stationary distribution if and only if all of its states are positive recurrent. Moreover, in that case, the stationary distribution is unique.

Note that there is no assumption on the starting distribution; the chain converges to the stationary distribution regardless of where it begins. In our case we have a finite irreducible Markov chain. We can then use the following well-known fact about Markov chains (see, e.g. [16]).

**FACT 2.** In a finite irreducible Markov chain all of its states are positive recurrent.

As a consequence of Fact 1 and Fact 2, there exists a unique stationary distribution regardless of how the nodes were deployed initially. Thus, we obtain a unique stationary distribution that can be computed efficiently. We conclude that any mobility model which defines time-homogeneous transition probabilities from one point to another point on the finite geographical model defines an irreducible, recurrent Markov chain which will have a unique stationary spatial node distribution. Once a stationary distribution is derived, a random geometric graph is obtained as in the proof of Theorem 2 and the edge parameter $p$ can be calculated as before. We thus obtain the following:

**Theorem 3.** Assuming an arbitrary distribution on the initial placement of the $n$ nodes, given the generalized geographic, mobility, and communication models of Section 3.1, at any discrete time $t \geq 0$, the communication graph $G(t) = (V, E(t))$ is a random geometric graph with $n$ nodes and edge probability $p$, where $p$ has an expression that is algorithmically computable (with efficient run time).

4. TWO EXAMPLE APPLICATIONS

We apply our modeling results to routing (specifically, to estimating the resiliency of known routing protocols in the presence of node compromise due to faults or attacks) and to misbehaviour detection (specifically, to estimating a critical trust parameter in known election-based protocols for the detection of misbehaving nodes).

**Survivable routing.** Realizing survivable routing (i.e., routing resilient to node faults or node captures by attackers) over arbitrary communication networks such as ad-hoc or vehicular networks (or networks where nodes have limited communication range) is an interesting research area with challenging and multi-dimensional problems. One approach has been advocated to solve this problem, based on perfectly secure message transmission protocols, as originally studied in [17] and later by a large number of follow-up papers in the cryptography and distributed computing research communities. In many of these approaches, a sender attempts to contact a receiver by using multiple communication paths so to achieve resilience against node faults or node captures by attackers. Thus, the number of available communication paths is itself a metric for the survivability of network routing and any operations based on it. In what follows, we observe that our results in the previous sections can be used to estimate an upper bound on this metric for vehicular networks that can be modeled using the previously defined geographic, mobility and communication models.

In a general graph modeling the network, the expected number of such paths is upper bounded by the expected number of neighbours of a node. Thus, the expected number of neighbours in a random geometric graph with parameters $n$ and $p$ can be used to estimate the survivability of routing in vehicular networks that follow our communication, geographic and mobility models. Moreover, since we can express $p$ as a function of known or easy to compute model parameters, we can do the same for this survivability metric. Towards that goal, we see that as a simple application of Theorem 2, the distribution of the number of neighbours $|N_i(t)|$ of node $i$ at time $t$ can be calculated as follows. Let $Z_i(t)$ be the set of points on the grid which are in the communication range of the node $i$ at time $t$; i.e.,

$$Z_i(t) = \{(x, y) : (x, y) \in t \land \|(x, y), (x, y)\| \leq r\},$$

and define $k_1 = (H_1 \cap Z_i(t)), k_2 = (H_2 \cap Z_i(t))$. Then the probability that a node $j$ is connected to node $i$ at time $t$ is the probability that node $j$’s position is in $k_1 \cup k_2$. Recall that Theorem 1 implies that each node position is a specific stationary stochastic process $S_j$, and thus, if all nodes are initially placed according to the stationary distributions $S_1 = \cdots = S_n$, we have that the probability that a node $j$ is connected to node $i$ at time $t$ is $p = |k_1|/P + 2|k_2|/P$, which has been previously computed in a closed form expression. Thus, the probability distribution $g(|N_i(t)|)$ of $i$’s number of neighbours $|N_i(t)|$ at time $t$ is binomial with parameters $n - 1$ and $p$; i.e.,

$$g(|N_i(t)|) = B \left( n - 1, \frac{|k_1|}{P} + \frac{2|k_2|}{P} \right).$$

In particular, we note that the expected value and variance of $|N_i(t)|$ are easily computable as $(n-1)p$ and $(n-1)p(1-p)$. 
Security of misbehaviour detection. As discussed earlier, connectivity of the network is closely related to the number of neighbours of a node [19]. A higher value of the expected number of neighbours of a node increases network connectivity. Moreover, to ensure that a node captured by an adversary is quickly and reliably detected, it has been suggested that election protocols among the neighbours are used (see, e.g., [20] and references therein). As the reliability of results output by these protocols typically depends on the assumption that the majority of participants is honest, it is of interest to evaluate conditions under which the number of the adversary’s neighbours is sufficiently large. In what follows, we answer the following question: how many mobile infrastructure vehicles (e.g., police cars) should be added to the network to increase the number of neighbours of a node such that the connectivity and security of vehicular communications is significantly improved?

Suppose that we require at least $\gamma$ neighbours for each node to achieve full connectivity of the network and quick malicious user detection. Then, we have to find out how many infrastructure vehicles have to be added to the network such that the probability of at least $\gamma$ neighbours for a node is greater than a required threshold $\delta$. Suppose $\phi$ infrastructure vehicles have to be added to ensure $\gamma$ neighbours with a probability greater than $\delta$. Assuming the infrastructure vehicles have the same communication range as all other vehicles and that they also follow the same mobility model as the other vehicles, the condition can be formulated as:

$$1 - \sum_{i=1}^{\gamma-1} \binom{n}{i} p_i (1-p)^{(n+i-1)} > \delta,$$

where $p$ is the probability of an edge calculated in previous sections. It is possible that, the infrastructure vehicles have higher capabilities such as a larger communication range, denoted as $r_{\phi}$. Then the probability of finding an infrastructure vehicle in a node’s neighbourhood is referred to as $p_{\phi}$, and is higher than $p$. Specifically, $p_{\phi}$ can be calculated in exactly the same way as done for $p$, after setting the communication radius to $r_{\phi}$. Then the condition of finding $\geq \gamma$ neighbours of a node with probability $> \delta$, is formulated as:

$$1 - \sum_{i=0}^{\gamma-1} \left( \sum_{j=1}^{i} \binom{n}{j} p_j (1-p)^{(n-j)} + \left( \frac{\phi}{1-\gamma} \right) p_{\phi} (1-p_{\phi})^{(n-(i-j))} \right) > \delta.$$

Both above conditions can be readily evaluated given known or easy to estimate network parameters, including those computed in the previous sections.

5. CONCLUSIONS

Our synthetic modeling of vehicular networks as random geometric graphs satisfies a special concreteness property: the random geometric graph parameters can be computed (via a closed-form or efficiently computable expression) as a function of the network’s real-life parameter values. Our results may be of interest in further studies that model vehicular networks as random geometric graphs, as follows. If vehicular networks are believed to satisfy our modeling assumptions, one can: (1) analyse protocols or applications on the random geometric graph abstraction of the original network; (2) translate the analysis results to the original network using our closed-form or algorithmically computable expressions. As an example, we studied neighbour and connectivity questions that are related to important security issues, recently raised in vehicular network research initiatives (e.g., [1, 2]), as they are of critical importance to ensure the security of all conventional vehicular network applications, such as traffic safety, traffic awareness and commercial transactions.

6. REFERENCES