Analysis of Connectivity and Capacity in One-Dimensional Vehicle-to-Vehicle Networks

Sungoh Kwon, Yoora Kim, and Ness B. Shroff

***Abstract*—A vehicle-to-vehicle network is one type of mobile ad-hoc network. Due to mobility, the topology in a vehicle-to-vehicle network is time-varying, which complicates the analysis and evaluation of network performance. In this paper, we model the network as geometric elements of lines and points and analyze connectivity and capacity of the network using geometric probability. Under the assumption that n vehicles randomly arrive with a Poisson distribution, our analysis shows that the spatial distribution of vehicles within a given distance, D, is uniform and that the average number of vehicles to be fully connected is approximately $\frac{1}{a}\left(\log\frac{1}{a}+\log\log\frac{1}{a}\right)$ for $a=\frac{R_T}{D}$, where R_T is the maximum transmission range of a vehicle. When a random access scheme is adopted, only $\frac{1}{2}(1-e^{-2})n$ of links comprised of two adjacent nodes are simultaneously activated, on average, so the expected network capacity increases in a way linearly proportional to $\frac{1}{2}(1-e^{-2})$ as the number of vehicles increases. Through numerical studies and simulations, we verify the efficacy of our analytical results.**

***Index Terms*—Vehicle-to-vehicle communications, geometric probability, performance analysis, connectivity, capacity**

I. Introduction

Vehicular networks have been studied extensively for application in intelligent transportation systems (ITS) for safety warnings [1], [2] as well as for data communications [3], [4]. In such networks, communications can be categorized into two modes: vehicle-to-infrastructure (V2I) communications and vehicle-to-vehicle (V2V) communications, as shown in Fig. 1. Vehicles communicate with roadside units (RSUs) to access backbone networks, referred to as V2I communications. Through V2I communications, ITS servers collect traffic data from vehicles and send traffic information to vehicles.

Between RSUs, not all vehicles can communicate with the RSUs in a single hop due to their limited transmission range. As a result, vehicles communicate with each other to exchange information via V2V communications. Since such communications do not have an infrastructure, data are delivered from one place to another in a multi-hop manner. For such vehicular networks, IEEE 802.11p, also known as wireless access in vehicular environment (WAVE), was proposed [5]. Since IEEE 802.11p is grounded in IEEE 802.11, the physical medium is shared with other communication nodes, and the carrier sense multiple access/collision avoidance (CSMA/CA) scheme is employed as a random access protocol.

S. Kwon is with the School of Electrical Engineering, University of Ulsan, Ulsan, Korea (email: sungoh@ulsan.ac.kr, nj0324@gmail.com).

Y. Kim is with the Department of Mathematics, University of Ulsan, Ulsan, Korea (Corresponding author, email: yrkim@ulsan.ac.kr).

N. B. Shroff is with the Department of Electrical and Computer Engineering and the Department of Computer Science and Engineering, The Ohio State University, OH, U.S.A. (email: shroff.11@osu.edu).

This research has been supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2015R1A1A1A05001069) and the Ministry of Education (NRF-2014R1A1A2057793); and by grants from the Army Research Foundation MURI W911NF-12-1-0385, the Defense Thrust Reduction Agency (DTRA) HDTRA1-14-1-0058, the Office of Naval Research N00014-15-1-2166, and the NSF CNS-1421576 and CNS-1409336.

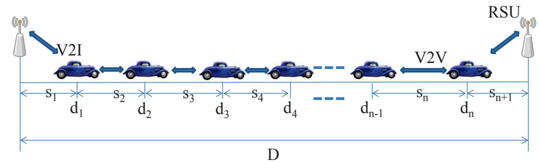


Fig. 1. A vehicular network consisting of n vehicles and two RSUs

Analyzing the performance of such networks involves non-trivial challenges due to the following two factors: random positions of vehicles and random links between vehicles. Temporal information on vehicles, such as inter-arrival times and velocities, can be measured at RSUs, but the spatial information on the vehicles between two adjacent RSUs is unknown, which is directly related to network connectivity. Due to vehicle mobility, vehicles are randomly located between RSUs, and the network topology changes with time. Hence, inter-vehicle distances may be greater than the transmission range, resulting in a *disconnected* vehicular network. Consequently, successful end-to-end communications in vehicular ad hoc networks (VANETs) depend on network connectivity, which impacts the main performance measures of a vehicular network, such as capacity and delay.

Even if all inter-vehicle distances are less than the transmission range (i.e., the vehicular network is topologically *fully connected*), data cannot be propagated immediately from source node to destination node in the network via the vehicles located in between them. A path from source node to destination node is composed of links, each of which is formed by an ordered pair of nodes (sender and receiver). Due to the random access scheme adopted by IEEE 802.11p, only a subset of links in the path is allowed to simultaneously activate. The random links as well as the network connectivity affect the throughput capacity and delay of the network.

In this paper, we consider a one-dimensional VANET formed on a unidirectional highway, addressing the following problems for performance analysis:

1) the spatial distribution of vehicles on the road between two adjacent RSUs
2) the geometric properties of the vehicular ad-hoc network
3) the probability that the vehicular ad-hoc network is fully

connected
4) the capacity of the vehicular ad-hoc network

In previous work, various aspects of connectivity in one-dimensional VANETs have been studied [6]–[12]. The connectivity of vehicular ad-hoc networks was theoretically analyzed in [7]–[12], but was studied by computer simulations in [6]. In [7], the authors examined the connectivity characteristics related to clusters[1] (such as average intra-cluster spacing and average cluster size[2]) in sparse one-way VANETs using mobility patterns extracted from real-world empirical data. In [8], the authors presented a comprehensive mobility model by considering the arrival and departure of nodes at predefined entry and exit points along a highway, and studied the average cluster size and the probability that the nodes form a single cluster. In [9], the authors studied a way to improve the connectivity in a VANET by adding mobile base-stations, and analyzed the average connectivity distance and the average cluster size. In [10], the author developed connectivity probability using a geometry-assisted analytical method. In [11], the authors introduced an equivalent speed to account for various speeds across vehicles for use in connectivity analysis. In [12], the connectivity of one-dimensional vehicular networks was analyzed with consideration of the fading channel.

While there have been extensive studies on connectivity in VANETs, the problem of characterizing network topology when the VANET is not fully connected has been under-explored. Specifically, statistics on the number of disconnected links has not been well understood, and as of yet, there is no analytic solution for these statistics. Whenever a VANET is not fully connected, the number of disconnected links can be different. That is, a disconnected VANET can have various degrees of disconnection or fragmentation. Such statistics on the number of disconnected links is important information in the design of a VANET that can provide reliable communications for vehicles. Recently, the use of Long Term Evolution (LTE) to support a VANET is under investigation by standardization groups [13]. Detailed knowledge on the degree of network disconnection also enables us to precisely estimate the required number of cluster heads or mobile relays that bridge vehicles in a VANET and eNodeBs under LTE [14]. This estimation in turn can help to allocate network resources in a more elaborate way for potential VANET–LTE interworking.

In this paper, we present an exact-form analysis of the number of disconnected links in a VANET in a unidirectional highway environment. The main techniques in the previous work on connectivity analysis in VANETs rely on queueing theory or geometric-assisted analytical models. These techniques have mostly been developed to analyze the *connection probability* of a VANET, which is equivalent to the probability of having no disconnected links. Since the topology of a VANET is spatiotemporally correlated and highly dynamic in nature, analyzing the entire spectrum of the number of disconnected links requires handling a huge dimension of diversity across all random links. To handle this diversity, we model the network as geometric elements of lines and points, and we utilize another mathematical technique based on geometric probability and finite difference calculus. This technique enables us to fully characterize the number of disconnected links in terms of probability distribution, which includes the connection probability as one case. In detail, previous works have provided only the probability that there are no disconnected links in the network, whereas our analysis provides the probability distribution that the network has k disconnected links for all values of k, including $k = 0$. This result translates into the probability that the network is composed of $(k+1)$ clusters for each of $k \in \{0, 1, \ldots\}$, from which we can obtain the average number of clusters.

Another important metric concerned with connectivity is traffic density (in vehicles per kilometer per lane) [15]. As the traffic density increases, the vehicular network is more likely to be fully connected. In this paper, we define *critical network size* to be the number of vehicles with which the vehicular network begins to be fully connected, and below which a link is disconnected. A similar notion can be found in percolation theory (the so-called percolation threshold [16], [17]) if we consider a vehicle as a circle where the radius equals the vehicle's transmission range. The critical network size represents a phase transition point in a VANET, since the connectivity of the VANET changes completely from a disconnected phase to a connected phase. In this paper, we present both the exact and asymptotic formulas for the average critical network size, and we analyze its behavior for various system parameters.

Our contributions in this paper are summarized as follows.

1) Based on the assumption that the arrival process of vehicles is Poisson, we show that the spatial distribution of vehicles between two adjacent RSUs is uniform. The average number of vehicles between two adjacent RSUs is linearly proportional to the vehicle arrival rate multiplied by $\mathrm{E}[V^{-1}]$, where V is a random variable denoting the vehicle velocity.
2) We analyze the connectivity of a VANET between two adjacent RSUs in a unidirectional highway environment. In particular, we first derive a closed-form expression for the probability distribution of the number of disconnected links. Using the distribution, we obtain the average number of clusters, as well as the probability that the VANET is fully connected. Our analysis shows the relationship between connectivity and the system parameters such as vehicle arrival rate, transmission range, and vehicle velocities.
3) We show that the average critical network size is approximately $\frac{1}{a}\left(\log\frac{1}{a} + \log\log\frac{1}{a}\right)$, where a is the ratio of the maximum transmission range of a vehicle to the distance between two adjacent RSUs.
4) We investigate the capacity of a one-dimensional VANET. We show that when network capacity is defined as the amount of data delivered in a network during a unit of time [18]–[20], the network capacity is exactly

[1]Throughout this paper, a cluster refers to a set of nodes formed as follows. Any two adjacent nodes belong to the same cluster if their inter-node distance is shorter than a given transmission range; otherwise, the two nodes belong to different clusters. Accordingly, a VANET is composed of $k \in \{1, 2, \ldots\}$ clusters when there are $(k-1)$ disconnected links in the network.

[2]The average cluster size denotes the expected number of nodes within a cluster.

expressed as $\frac{1}{2}(1-e^{-2})n$ in an asymptotic regime, in contrast to $\Theta\left(\sqrt{\frac{n}{\log n}}\right)$ in two-dimensional random ad-hoc networks [18], and $\Theta(n)$ in one- or two-dimensional mobile ad-hoc networks [19], [20], where n is the number of nodes.

The rest of this paper is organized as follows. In Section II, we describe the system model. In Section III, we study the spatial distribution of vehicles based on temporal distribution and velocity information of vehicles. In Sections IV and V, we analyze the connectivity and capacity of the network. In Section VI, we present numerical studies. Finally, we conclude the paper in Section VII.

II. System Model and Assumptions

In this paper, we consider a segment of a unidirectional multi-lane highway that consists of two adjacent RSUs and $N(t)$ vehicles between them at time t. Each vehicle is equipped with a communication device that has a maximum transmission range, R_T, and can control its transmission power. The RSUs are spaced apart at distance D, which is much longer than transmission range R_T. Thus, vehicles between the two RSUs can communicate with each other only in a multi-hop fashion without relying on a wired infrastructure.

In practice, the transmission range of an RSU is wider than that of a vehicle. However, if a vehicle is located farther away from an RSU than its transmission range, then the vehicle cannot bidirectionally communicate with the RSU. Hence, we assume that the RSUs also have maximum transmission range R_T, enabling bidirectional communication between a vehicle and an RSU.

The topology of the VANET changes with time due to vehicle mobility. The vehicle mobility model used in this paper can be described in terms of vehicle arrival and velocity as follows. With regard to vehicle arrival, we assume that each vehicle arrives at the first RSU according to a Poisson process, as assumed in previous works [8]–[12], [21]. Empirical studies from real freeway traffic data have shown that vehicle arrivals can be accurately approximated by a Poisson process, especially for a sparse traffic environment, such as night time periods and early mornings [7]. With regard to vehicle velocities, we assume that the speed of a vehicle is a random variable that is independent for each vehicle and follows a *general* probability distribution. Vehicles are assumed to freely overtake slower vehicles so that their velocities are not hindered by other vehicles. This assumption practically holds in sparse situations [21] or in free-flow phase in the traffic theory language [9], [15]. For example, in the case of multi-lanes in rural areas, vehicle speeds are likely to be unaffected by the traffic. Hence, we assume that each vehicle drives at a constant speed of its own, which is chosen randomly, until reaching the second RSU.

Due to the time-varying topology, $\{N(t)\}_{t\geq 0}$ is a random process. For a fixed time t, suppose that $n\,(\in \mathbb{N}\cup\{0\})$ vehicles are located between the two RSUs, i.e., $N(t)=n$. Throughout this paper, we count vehicles and RSUs together as nodes and label these $n+2$ nodes as Node $k\,(k=0,1,\ldots,n+1)$ in the vehicle driving direction, i.e., the first and the second RSUs are Nodes 0 and $n+1$, respectively, and the n vehicles are Nodes 1 to n. Let $d_k(t)\,(\in \mathbb{R})$ be the location of Node k along the direction of the highway[3] at time t. Without loss of generality, we set $d_0(t)=0$ and $d_{n+1}(t)=D$ for all t. Then,

$$0=d_0(t)<d_1(t)<\ldots<d_n(t)<d_{n+1}(t)=D. \quad (1)$$

The spatial distribution of the n ordered random vector $(d_1(t),d_2(t),\ldots,d_n(t))$ determines the state of each link between nodes and overall network connectivity. The link between Nodes $k-1$ and k at time t is topologically either *connected* or *disconnected*, and becomes connected if the inter-node distance at time t is shorter than the maximum transmission range, as follows:

$$s_k(t)\triangleq d_k(t)-d_{k-1}(t)\leq R_T.$$

In addition, the vehicular network is topologically *fully connected* at time t if no link is disconnected at time t, i.e., $s_k(t)\leq R_T$ for all $k=1,2,\ldots,n+1$. For such topological connectivity, we ignore interference from other transmissions, which may affect the physical link connectivity.

A random access scheme is employed to establish a communication link between two nodes. We assume that nodes transfer data only to adjacent nodes, and each node can form at most a single link at a time. For example, if Node k communicates with Node $k+1$ at time t, then Node $k+2$ cannot form a link with Node $k+1$ at that time. A node can generally reach any node in its transmission range. However, in such a case, intermediate nodes between the two communicating nodes (the sender and the receiver) cannot communicate with each other due to collision at receivers, and as a result, the network capacity decreases [18], [23], [24]. Moreover, the greater transmission range induces more collisions at other communication links, and decreases network capacity. Hence, in this paper, a node with power control is assumed to establish a communication link with an adjacent node. We define that link between Nodes k and $k+1$ as *active* at time t if Node k communicates with Node $k+1$ at time t. Otherwise, the link between Nodes k and $k+1$ is *idle*, even though Node k is topologically connected to Node $k+1$ at time t. Since IEEE 802.11p provides multiple channels, nodes can communicate with their neighboring nodes without the hidden node problem or the exposed node problem [25]–[27].

Under the above assumptions, in this paper, we first analyze the spatial distribution of vehicles on the road between two adjacent RSUs. Based on this spatial distribution, we next study the connectivity of a vehicular network from various aspects. We then derive the capacity of a vehicular network employing a random access scheme.

III. Distribution of Vehicles

In this section, we determine the spatial distribution of vehicles on the road between two adjacent RSUs at a given time,

[3] A more precise expression for the location of Node k is $(d_k(t),y_k(t))\in\mathbb{R}^2$, where $y_k(t)$ is the signed distance from a fixed point (e.g., an RSU) to Node k along a line perpendicular to the highway. It is shown that the effect of $y_k(t)$ on connectivity can be considered negligible in multi-lane highways for most practical cases, since the transmission range is $R_T \gg 3.6\ m$ that is the standard highway's lane width [15], [22]. Hence, we express the location of Node k using $d_k(t)$.

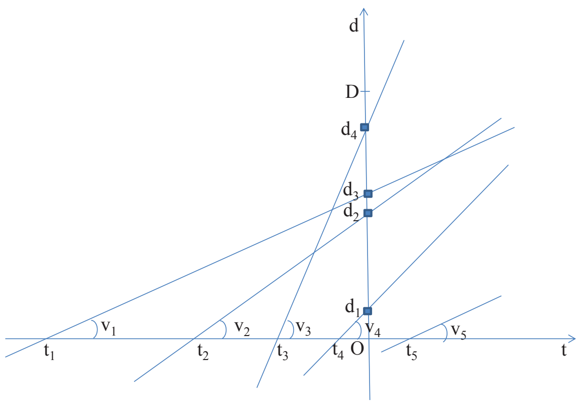


Fig. 2. Modeling of a vehicular network using geometry

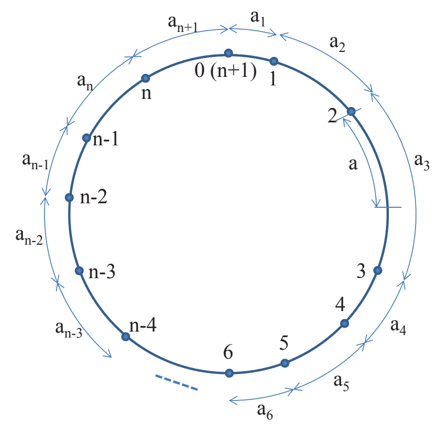


Fig. 3. Network connectivity when there are $N(t) = n$ vehicles between two adjacent RSUs at time t

based on the temporal distribution and velocity information of vehicles gathered at the RSUs. The trajectory of a vehicle created using time and space coordinates can be modeled as a straight line in a plane in $\mathbb{R}^2$ if the vehicle drives at a constant speed and never leaves the road, as shown in Fig. 2. In the figure, the horizontal axis and the vertical axis represent, respectively, time t and distance d of the vehicle from the first RSU.

The trajectory of vehicle i can be parameterized by (V_i, t_i) and is expressed as

$$d = V_i(t - t_i).$$

Here, the time intercept t_i on the t-axis denotes the time when vehicle i passes the first RSU, and V_i is the velocity of vehicle i. Since the arrival epochs $\{t_i\}_{i\in\mathbb{N}}$ form a Poisson process, and the velocities $\{V_i\}_{i\in\mathbb{N}}$ are independent and identically distributed across i, it follows that the spatial distribution of vehicles at any given time t is Poisson [28], [29]. Also, its density, denoted by λ_d, is determined as

$$\lambda_d = \lambda \int_0^\infty \frac{1}{v} dF_V(v) = \lambda \mathrm{E}[V^{-1}], \tag{2}$$

where λ is the arrival rate of the vehicles, and $F_V(\cdot)$ is the distribution function of velocity V. It means that the spatial density λ_d of vehicles between two RSUs is the measured temporal density λ scaled by $\mathrm{E}[V^{-1}]$. As a consequence, the average number of vehicles between the RSUs is $\mathrm{E}[N(t)] = \lambda_d D$ for any time t, and the probability mass function of $N(t)$ is obtained by

$$\mathrm{P}(N(t) = n) = e^{-\lambda_d D}\frac{(\lambda_d D)^n}{n!}, \qquad n = 0, 1, \ldots. \tag{3}$$

Due to stationary and independent increments inherent in a Poisson process, the joint distribution of $(d_1(t), \ldots, d_n(t))$ when $N(t) = n$ can be described by a uniform distribution on the interval $(0, D)$ [30], as in Lemma 1.

Lemma 1. *Given that $N(t) = n$, the joint density function of $(d_1(t), d_2(t), \ldots, d_n(t))$ in (1) is*

$$f(x_1, x_2, \ldots, x_n) = \frac{n!}{D^n},$$

where $0 < x_1 < x_2 < \ldots < x_n < D$.

Proof: Refer to [30]. ■

The result in Lemma 1 implies that $d_k(t)$ has the same distribution as the kth smallest value among n independent random variables *uniformly* distributed on the interval $(0, D)$. Therefore, if we are given the information on the number of vehicles between the RSUs at time t (say, $N(t) = n$), then we can infer that the n vehicles are uniformly distributed on the road.

IV. Connectivity of a VANET

Based on spatial distribution analysis, we next analyze the connectivity of a VANET. Specifically, we derive the following four metrics in sequence: (i) the probability distribution of the number of disconnected links, (ii) the average number of clusters, (iii) the probability that the VANET is fully connected, and (iv) the average critical network size.

A. Probability distribution of the number of disconnected links

Since RSUs are evenly located, and the segments between two adjacent RSUs have identical statistics, we convert a line segment into a circle to remove edge effects as in [19], [20].[4] The detailed procedure is as follows. We first normalize the line segment in Fig. 1 by length D. We next transform the scaled line segment into a circle by attaching the first and the second RSUs, as shown in Fig. 3. We then denote the RSU at the origin in Fig. 2 by zero on the circle and assign numbers from Node 0 clockwise.

Suppose that $n\,(\in \mathbb{N} \cup \{0\})$ vehicles are located between the two RSUs at time t, i.e., $N(t) = n$. It then follows from Lemma 1 that the vehicles are located randomly on the circumference of the circle with a uniform distribution. In addition, there are $n+1$ arcs on the circle, and the arc between Nodes $k-1$ and k (i.e., the kth arc) on the circle corresponds

[4] We are inspired by the modelling in [31], and our analysis follows the methodology in there.

to the kth link in the VANET and has the normalized arc length $a_k(t)$, defined as

$$a_k(t) \triangleq \frac{s_k(t)}{D} = \frac{d_k(t) - d_{k-1}(t)}{D}, \qquad k = 1, 2, \ldots, n+1.$$

Hence, the sum of all $a_k(t)$s is one, i.e., $\sum_{k=1}^{n+1} a_k(t) = 1$.

Let a be the ratio of transmission range R_T to distance D between two adjacent RSUs, i.e.,

$$\text{(Normalized transmission range)} \quad a \triangleq \frac{R_T}{D}.$$

In practice, the transmission ranges are not constant, but randomly distributed. Since the statistical properties of a constant transmission range case and a random transmission range case are similar when the number of nodes n is large [32], we study the connectivity of a VANET under the assumption that the transmission range is constant. If arc length $a_k(t)$ is greater than the normalized transmission range a, then Nodes $k-1$ and k are unable to communicate with each other. If such an event occurs, we say that the kth arc has a gap, which is also equivalent to having a topologically disconnected link between Nodes $k-1$ and k in the vehicular network. When no gap occurs in any arc on the circle, the VANET is fully connected.

To facilitate our analysis, we first compute the probability $g_n(i)\,(i = 1, 2, \ldots, n+1)$ that gaps occur in i specified arcs (namely, k_1th, k_2th, $\ldots, k_i$th arcs), whatever happens elsewhere; i.e.,

$$g_n(i) \triangleq \mathrm{P}(a_k(t) > a, k = k_1, k_2, \ldots, k_i \,|\, N(t) = n).$$

Due to circular symmetry, $g_n(1)$ is identical to the probability that the first arc has a gap, which is $(1-a)^n$ since n vehicles are uniformly distributed on the circle by Lemma 1. Applying a similar argument, the probability $g_n(i)$ is then derived as

$$g_n(i) = \begin{cases} (1 - ia)^n & \text{if } \; i \le \lfloor a^{-1} \rfloor, \\ 0 & \text{if } \; i > \lfloor a^{-1} \rfloor, \end{cases} \tag{4}$$

where $\lfloor x \rfloor$ is the largest integer not greater than x.

We next compute the probability $\hat{g}_n(i)\,(i = 0, 1, \ldots, n+1)$ that exactly i gaps occur on the circle, wherever they happen, i.e.,

$$\hat{g}_n(i) \triangleq \mathrm{P}\left(\sum_{k=1}^{n+1} 1_{\{a_k(t) > a\}} = i \,\middle|\, N(t) = n \right), \tag{5}$$

where $1_{\{\cdot\}}$ is an indicator function. From the calculus of finite differences and the results in [33] and [34], the probability $\hat{g}_n(i)$ can be obtained by

$$\hat{g}_n(i) = \binom{n+1}{i} \sum_{j=0}^{n+1-i} \binom{n+1-i}{j} (-1)^j g_n(i+j), \tag{6}$$

where $\binom{n}{j}$ is the number of ways of choosing j unordered outcomes from n possibilities. In this paper, we adopt the convention that $\binom{n}{j} = 0$ when $j > n$ or $j < 0$. Combining equations (4) and (6), we have the following theorem on the distribution of the number of disconnected links.

Theorem 1. *Given that $N(t) = n$, the probability that there are i disconnected links in the network is*

$$\hat{g}_n(i) = \begin{cases} \binom{n+1}{i} \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j (1-(i+j)a)^n, \\ \qquad \textit{if} \;\; i \le \lfloor a^{-1} \rfloor, \\ 0, \qquad \textit{if} \;\; i > \lfloor a^{-1} \rfloor, \end{cases} \tag{7}$$

where $i = 0, 1, \ldots, n+1$.

Proof: From (4), we have

$$g_n(i+j) = \begin{cases} (1 - (i+j)a)^n & \text{if } \; j \le \lfloor a^{-1} \rfloor - i, \\ 0 & \text{if } \; j > \lfloor a^{-1} \rfloor - i. \end{cases} \tag{8}$$

Suppose first that $\lfloor a^{-1} \rfloor - i < 0$. Then, $g_n(i+j) = 0$ for all $j \ge 0$, and thus, we have $\hat{g}_n(i) = 0$ by (6). Suppose next that $\lfloor a^{-1} \rfloor - i \ge 0$. Then, $g_n(i+j) = 0$ for all $j > \lfloor a^{-1} \rfloor - i \ge 0$. Hence, substituting the first case in (8) into (6) gives

$$\begin{aligned} \hat{g}_n(i) &= \binom{n+1}{i} \sum_{j=0}^{\min(n+1-i, \lfloor a^{-1} \rfloor - i)} \binom{n+1-i}{j} (-1)^j (1-(i+j)a)^n \\ &= \binom{n+1}{i} \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j \, (1-(i+j)a)^n, \end{aligned}$$

where the second equality follows, since $\binom{n+1-i}{j} = 0$ for $j > n+1-i$. This completes the proof. ∎

B. Expected number of clusters

A cluster, which is also referred to as a platoon, denotes a set of nodes in which the distance between any two adjacent nodes is less than the transmission range. The number of clusters would be an interesting parameter in a certain type of a VANET. Clustering technology provides efficient and reliable communications in a VANET, as well as an integration of VANET and cellular networks [3], [4], by communicating with neighboring nodes, referred to as cluster members. In such cluster-based networks, cluster heads are elected to manage both intra-cluster and inter-cluster communications.

Let $C(t)$ be a random variable denoting the number of clusters in the network at time t. According to the discussion above, a VANET forms k clusters when there are $(k-1)$ disconnected links in the network. Hence, applying the result in Theorem 1, we can derive the average number of clusters in the network, as shown in Theorem 2.

Theorem 2. *Given that $N(t) = n$, the expected number of clusters in the network is*

$$\mathrm{E}[C(t) \,|\, N(t) = n] = \sum_{i=0}^{\lfloor a^{-1} \rfloor} (i+1) \binom{n+1}{i} \cdot \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j \, (1-(i+j)a)^n. \tag{9}$$

Since $N(t)$ follows a Poisson distribution, the expected number of clusters in the steady state becomes

$$\mathrm{E}[C(t)] = e^{-\lambda_d D} \sum_{n=0}^{\infty} \sum_{i=0}^{\lfloor a^{-1} \rfloor} (i+1) \binom{n+1}{i} \cdot \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j \frac{\{\lambda_d D(1-(i+j)a)\}^n}{n!}. \quad (10)$$

Proof: Given $N(t) = n$, we have $C(t) = k$ with probability $\hat{g}_n(k-1)$ for $k = 1, 2, \ldots, n+2$. Hence,

$$\begin{aligned}
&\mathrm{E}[C(t) \,|\, N(t) = n] \\
&= \sum_{k=1}^{n+2} k \cdot \mathrm{P}(C(t) = k \,|\, N(t) = n) \\
&= \sum_{k=1}^{n+2} k \cdot \hat{g}_n(k-1) \\
&= \sum_{i=0}^{n+1} (i+1) \cdot \hat{g}_n(i) \\
&= \sum_{i=0}^{\min(n+1, \lfloor a^{-1} \rfloor)} (i+1) \binom{n+1}{i} \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j (1-(i+j)a)^n \\
&= \sum_{i=0}^{\lfloor a^{-1} \rfloor} (i+1) \binom{n+1}{i} \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j (1-(i+j)a)^n,
\end{aligned}$$

where the last equality follows, since $\binom{n+1}{i} = 0$ for $i > n+1$.

By the law of total probability, the expected number of clusters in the steady state is then derived as

$$\begin{aligned}
\mathrm{E}[C(t)] &= \sum_{n=0}^{\infty} \mathrm{E}[C(t) \,|\, N(t) = n] \mathrm{P}(N(t) = n) \\
&= e^{-\lambda_d D} \sum_{n=0}^{\infty} \sum_{i=0}^{\lfloor a^{-1} \rfloor} (i+1) \binom{n+1}{i} \\
&\quad \cdot \sum_{j=0}^{\lfloor a^{-1} \rfloor - i} \binom{n+1-i}{j} (-1)^j \frac{\{\lambda_d D(1-(i+j)a)\}^n}{n!},
\end{aligned}$$

where the second equality follows from (3). This completes the proof. ∎

The above theorem can answer the question of how many cluster heads exist in the VANET, when clusters are formed based on connectivity and each cluster has a cluster head. While previous works studied efficient clustering algorithms and their performance, our analysis provides the number of cluster heads, which was not studied in the previous works.

C. Connection probability of a VANET

When the network forms only a single cluster, the VANET is fully connected. Hence, using Lemma 1 and Theorem 1, we can derive the connection probability of a VANET, as summarized in Theorem 3.

Theorem 3. *Given that $N(t) = n$, the network becomes fully connected with the following probability:*

$$\begin{aligned}
&\mathrm{P}(\text{VANET is connected} \,|\, N(t) = n) \\
&= \textstyle\sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{n+1}{j} (-1)^j (1-ja)^n. \quad (11)
\end{aligned}$$

Since $N(t)$ follows a Poisson distribution, the connection probability of a VANET in the steady state is thus obtained by

$$\begin{aligned}
&\mathrm{P}(\text{VANET is connected}) \\
&= e^{-\lambda_d D} \sum_{n=0}^{\infty} \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{n+1}{j} (-1)^j (1-ja)^n \frac{(\lambda_d D)^n}{n!}. \quad (12)
\end{aligned}$$

Proof: According to the definition of $\hat{g}_n(\cdot)$ in (5), we have

$$\mathrm{P}(\text{VANET is connected} \,|\, N(t) = n) = \hat{g}_n(0).$$

Hence, by Theorem 1, we obtain

$$\hat{g}_n(0) = \begin{cases} \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{n+1}{j} (-1)^j \, (1-ja)^n & \text{if } \; 0 \le \lfloor a^{-1} \rfloor, \\ 0 & \text{if } \; 0 > \lfloor a^{-1} \rfloor. \end{cases}$$

Since $0 < a \le 1$, we have $\lfloor a^{-1} \rfloor \ge 0$. Thus,

$$\begin{aligned}
&\mathrm{P}(\text{VANET is connected} \,|\, N(t) = n) \\
&= \textstyle\sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{n+1}{j} (-1)^j (1-ja)^n.
\end{aligned}$$

This proves (11). By the law of total probability, the connection probability of a VANET in the steady state is then derived as

$$\begin{aligned}
&\mathrm{P}(\text{VANET is connected}) \\
&= \sum_{n=0}^{\infty} \mathrm{P}(\text{VANET is connected} \,|\, N(t) = n) \mathrm{P}(N(t) = n) \\
&= e^{-\lambda_d D} \sum_{n=0}^{\infty} \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{n+1}{j} (-1)^j (1-ja)^n \frac{(\lambda_d D)^n}{n!},
\end{aligned}$$

where the second equality follows from (3). This completes the proof. ∎

We can interpret the probability (12) in Theorem 3 as the percentage of time that the network is fully connected. For applications, if there exists a data center gathering traffic statistics, such as the average of vehicle arrivals, the long-term average of VANET connectivity can be computed from (12) in Theorem 3. Also, when information on the number of vehicles on the road is available in real time, (11) in Theorem 3 becomes useful for predicting the connectivity of the VANET over time.

D. Critical network size

In this section, we derive both the exact and asymptotic formulas for the average critical network size, i.e., the expected number vehicles required for a VANET to be fully connected.

Let N_a be a random variable denoting the critical network size, provided that the normalized transmission range a is used. Note that if the event $\{N_a \le m\}$ occurs, then a VANET consisting of m vehicles is fully connected. On the other hand, if a VANET with m vehicles is fully connected, then $\{N_a \leq m\}$. Hence,

$$\mathrm{P}(N_a \leq m) = \mathrm{P}(\text{VANET is connected} \,|\, N(t) = m).$$

From (11) in Theorem 3, we also have

$$\mathrm{P}(N_a \leq m) = \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{m+1}{j} (-1)^j (1 - ja)^m. \tag{13}$$

Based on the distribution function in (13), we present the closed-form formula for $\mathrm{E}[N_a]$ and its asymptotics in the following theorem.

Theorem 4. *The average critical network size of a VANET that uses normalized transmission range a is given by*

$$\mathrm{E}[N_a] = \sum_{m=1}^{\infty} \sum_{j=1}^{\lfloor a^{-1} \rfloor} \binom{m}{j} (-1)^{j-1} (1 - ja)^{m-1} \tag{14}$$

$$= \frac{1}{a} \left\{ \log \frac{1}{a} + \log\log \frac{1}{a} + \gamma + o(1) \right\}, \tag{15}$$

where γ is Euler's constant,[5] *and $o(1)$*[6] *is a negligibly small number compared to a.*

Proof: Since N_a takes on only non-negative integer values, we obtain the expectation $\mathrm{E}[N_a]$ as

$$\begin{aligned} \mathrm{E}[N_a] &= \sum_{m=1}^{\infty} \mathrm{P}(N_a \geq m) \\ &= \sum_{m=1}^{\infty} (1 - \mathrm{P}(N_a \leq m-1)) \\ &= \sum_{m=1}^{\infty} \left\{ 1 - \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{m}{j} (-1)^j (1 - ja)^{m-1} \right\} \\ &= \sum_{m=1}^{\infty} \sum_{j=1}^{\lfloor a^{-1} \rfloor} \binom{m}{j} (-1)^{j-1} (1 - ja)^{m-1}, \end{aligned} \tag{16}$$

which proves (14). Now, we investigate the asymptotic behavior of $\mathrm{E}[N_a]$ as normalized transmission range a approaches zero. By the Taylor series expansion, we have $1 - ja \approx e^{-ja}$ for small a. Hence, the summation on the right-hand side of the third equality in (16) becomes

$$\begin{aligned} \sum_{j=0}^{\lfloor a^{-1} \rfloor} \binom{m}{j} (-1)^j (1 - ja)^{m-1} &\approx \sum_{j=0}^{m} \binom{m}{j} (-1)^j e^{-jam} \quad (a \to 0) \\ &= \sum_{j=0}^{m} \binom{m}{j} (-e^{-am})^j \\ &= (1 - e^{-am})^m. \end{aligned} \tag{17}$$

[5]Euler's constant γ is defined as $\gamma = \lim_{n\to\infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \log(n) \right) \approx 0.57721$ [35].

[6]For two real functions $f(x)$ and $g(x)$, we use $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$ and $f(x) = O(g(x))$ if there exist constants c and $\hat{x} \in \mathbb{R}$ such that $|f(x)| \leq c|g(x)|$ for all $x \geq \hat{x}$. Hence, $o(1)$ implies a small number compared to a.

Fig. 4. Example of randomly activated links in a VANET

Applying the result in (17) to the third equality in (16) gives

$$\begin{aligned} \mathrm{E}[N_a] &\approx \sum_{m=1}^{\infty} \left\{ 1 - (1 - e^{-am})^m \right\} \\ &= \sum_{m \leq a^{-1}} \left\{ 1 - (1 - e^{-am})^m \right\} \\ &\quad + \int_{a^{-1}}^{\infty} \left\{ 1 - (1 - e^{-am})^m \right\} \mathrm{d}m + O(1) \\ &= O(a^{-1}) + a^{-1} \int_{1}^{\infty} \left\{ 1 - (1 - e^{-x})^{\frac{x}{a}} \right\} \mathrm{d}x, \end{aligned} \tag{18}$$

where change of variables with $am = x$ is used in the last equality. It then follows from [36, Equations (20), (21)] that

$$\begin{aligned} \int_{1}^{\log \frac{1}{a}} \left\{ 1 - (1 - e^{-x})^{\frac{x}{a}} \right\} \mathrm{d}x &= \log \frac{1}{a} - 1 + o(1), \\ \int_{\log \frac{1}{a}}^{\infty} \left\{ 1 - (1 - e^{-x})^{\frac{x}{a}} \right\} \mathrm{d}x &= \log\log \frac{1}{a} + \gamma + o(1). \end{aligned} \tag{19}$$

Combining (18) and (19), we have (15). ■

Applying Theorem 4, we can easily estimate the average number of vehicles required for a network connection under a given set of system parameters. For example, if the distance between adjacent RSUs is 2 km, and the maximum transmission range is 400 m (i.e., $a = 0.2$), then at least 14 vehicles on average must be moving on the road for the vehicular network to be fully connected (i.e., $\mathrm{E}[N_a] \approx 13.3$). In that case, the network is composed of 15 links, which are defined as a pair of two contiguous nodes.

V. Capacity Analysis of Random Access in a VANET

Since IEEE 802.11p employs CSMA/CA as a medium access control (MAC) protocol, links are randomly activated. As in Fig. 4, if Nodes 2 and 4 are communicating with Nodes 1 and 5, respectively, then Node 3 must be idle. Hence, among $N(t) + 1$ links, only a random portion of links participates in transmitting data in the VANET at time t, even though the VANET is fully connected at that moment.

We define the network capacity at time t as the summation of all transmitted data at time t. When the wireless bandwidth W is given, and the number of activated links is L at time t, the network capacity is expressed as WL. For simplicity, we assume that $W = 1$. Then, the expected number of links simultaneously activated by the random access, denoted by $\mathrm{E}[L]$, is directly related to network capacity.

In this section, we analyze $\mathrm{E}[L]$ as a function of $N(t) = n$ and study its asymptotic behavior as n approaches infinity. In order to focus on the random access effect, our analysis in this section is based on the regime $n \gg \mathrm{E}[N_a]$, so that the network is in a connected state. In this case, we have

$$\mathrm{E}[L|N(t)=n] = \frac{n+2-\mathrm{E}\big[N_{idle}|N(t)=n\big]}{2}, \tag{20}$$

where N_{idle} is the number of idle nodes that do not participate in communications. For example, if four vehicles are distributed between two RSUs, then the VANET consists of six nodes. If two pairs of nodes, Nodes 1 and 2 and Nodes 4 and 5, are communicating, then Nodes 0 and 3 are idle. Hence, the number of communicating links is two ($= \frac{6-2}{2}$).

To characterize the number of idle nodes, N_{idle}, we use the technique applied for solving a packing problem [30]. For $k = 0, 1, \ldots, N(t)+1$, we let

$$I_k(t) \triangleq \begin{cases} 1 & \text{if Node } k \text{ is idle at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the number of idle nodes can be expressed as $N_{idle} = I_0(t) + \ldots + I_{N(t)+1}(t)$. Hence,

$$\begin{aligned} \mathrm{E}[N_{idle}|N(t)=n] &= \mathrm{E}\left[\sum_{k=0}^{N(t)+1} I_k(t) \,\middle|\, N(t)=n\right] \\ &= \sum_{k=0}^{n+1} \mathrm{E}\big[I_k(t)\,\big|\,N(t)=n\big] \\ &= \sum_{k=0}^{n+1} p_{n+2}(k), \end{aligned} \tag{21}$$

where $p_{n+2}(k) \triangleq \mathrm{P}(I_k(t)=1\,|\,N(t)=n)$ denotes the probability that Node k is idle at time t, given that there are $N(t)+2 = n+2$ nodes in the network. To compute $p_{n+2}(k)$, we consider two contiguous segments:

$$\underbrace{0, 1, \cdots, k}_{k+1} \text{ and } \underbrace{k, k+1, \cdots, n, n+1}_{n+2-k}.$$

Node k will be idle if and only if the right-most node in the first segment and the left-most node in the second segment are both idle. Accordingly, if $\hat{p}_{m+2}\,(m = 0, 1, \ldots)$ is the probability that the end node in a VANET consisting of m vehicles is idle, i.e.,

$$\hat{p}_{m+2} \triangleq p_{m+2}(0) = p_{m+2}(m+1),$$

then $p_{n+2}(k)$ can be obtained from $\hat{p}_{m+2}$ as

$$p_{n+2}(k) = \begin{cases} \hat{p}_{n+2} & \text{if } k = 0, n+1, \\ \hat{p}_{k+1}\cdot\hat{p}_{n+2-k} & \text{if } k = 1, \ldots, n. \end{cases} \tag{22}$$

In the following, we derive a general expression for $\hat{p}_{m+2}$ for $m = 0, 1, \ldots$. To that end, we consider $m+2$ nodes distributed in a line, and suppose that link $L_{(i-1,i)}\,(i = 1, 2, \ldots, m+1)$ between Nodes $i-1$ and i is initially activated. Note the following: (i) this initial configuration is observed with probability $\frac{1}{m+1}$, since the random access scheme operates in a uniform manner[7]; (ii) if $i = m+1$ (i.e., link $L_{(m,m+1)}$ is activated), then Node $m+1$ cannot be idle trivially; (iii) if $i = m$ (i.e., link $L_{(m-1,m)}$ is activated), then Node $m+1$ is idle; and (iv) if $i \le m-1$, we separate the array into two independent segments:

$$\underbrace{0, 1, \cdots, i-2}_{i-1} \text{ and } \underbrace{i+1, \cdots, m, m+1}_{m+1-i}.$$

Then, Node $m+1$ is idle if the right-end node of the second segment consisting of $m+1-i$ nodes is idle. Combining (i)-(iv), we have

$$\begin{aligned} \hat{p}_{m+2} &= \frac{1}{m+1}\left(0 + 1 + \sum_{i=1}^{m-1}\hat{p}_{m+1-i}\right) \\ &= \frac{1+\hat{p}_2+\ldots+\hat{p}_m}{m+1}. \end{aligned} \tag{23}$$

We can rewrite (23) in recursive form as

$$\hat{p}_{m+2} - \hat{p}_{m+1} = -\frac{\hat{p}_{m+1}-\hat{p}_m}{m+1}.$$

Since $\hat{p}_2 = 0$ and $\hat{p}_3 = \frac{1}{2}$, we can solve for $\hat{p}_{m+2}$ as follows:

$$\hat{p}_{m+2} = \sum_{j=0}^{m+1}\frac{(-1)^j}{j!}, \quad m = 0, 1, \ldots. \tag{24}$$

We are now ready to derive the main result of this section. By substituting (22) and (24) into (21) and then using (20), we can obtain the network capacity, as stated in Theorem 5.

Theorem 5. *For $n \gg \mathrm{E}[N_a]$, the expected network capacity is approximately given by*

$$\mathrm{E}[L|N(t)=n] \approx \frac{1}{2}(n+2)(1-e^{-2}) - e^{-2}. \tag{25}$$

Proof: Substituting (24) into (22) gives

$$p_{n+2}(k) = \begin{cases} \sum_{j=0}^{n+1}\frac{(-1)^j}{j!} & \text{if } k = 0, n+1, \\ \left(\sum_{j=0}^{k}\frac{(-1)^j}{j!}\right)\left(\sum_{j=0}^{n+1-k}\frac{(-1)^j}{j!}\right) & \text{if } k = 1, \ldots, n. \end{cases} \tag{26}$$

Note that $p_{n+2}(0) = p_{n+2}(n) = 0$, because the first and the second term in (26) are zero for $k = 1, n$, respectively. Hence, from (21), we have

$$\mathrm{E}[N_{idle}|N(t)=n] = 2\sum_{j=0}^{n+1}\frac{(-1)^j}{j!} + \sum_{k=2}^{n-1}\left(\sum_{j=0}^{k}\frac{(-1)^j}{j!}\right)\left(\sum_{j=0}^{n+1-k}\frac{(-1)^j}{j!}\right).$$

Since $\lim_{n\to\infty}\sum_{j=0}^{n+1}\frac{(-1)^j}{j!} = e^{-1}$, the expectation $\mathrm{E}[N_{idle}|N(t)=n]$ as $n \to \infty$ behaves asymptotically as

$$\begin{aligned} \mathrm{E}[N_{idle}|N(t)=n] &\approx 2e^{-1} + (n-2)e^{-2} \\ &= (2e+n-2)e^{-2} \\ &\approx (n+4)e^{-2}. \end{aligned} \tag{27}$$

Therefore, by substituting (27) into (20), we have

$$\begin{aligned} \mathrm{E}[L|N(t)=n] &\approx \frac{n+2-(n+4)e^{-2}}{2} \\ &= \frac{1}{2}(n+2)(1-e^{-2}) - e^{-2}, \end{aligned}$$

[7]According to [37], the probability that a node transmits in a randomly chosen slot time is identical in the steady state.

which completes the proof. ■

The result in Theorem 5 indicates that, among $n+1$ possible links, only $\frac{1}{2}(n+2)(1-e^{-2})-e^{-2}$ links are simultaneously activated, i.e., approximately $\frac{n}{2}(1-e^{-2})$ for large n.

VI. Simulations

In this section, we examine the connectivity and capacity of a VANET for various system parameters through numerical studies and simulations. The simulation environments used in this section are as follows. We consider a segment of a unidirectional highway between two adjacent RSUs. The distance between the adjacent RSUs is set to $D = 3000$ m. The vehicle velocity is randomly chosen at between 80 km/h and 110 km/h. We assume that there are two lanes on the highway, and that vehicles maintain their velocities between the adjacent RSUs, even when overtaking other vehicles. The number of vehicle arrivals at the first RSU is modeled as a Poisson random variable, and the vehicle arrival rate λ is set to 30 vehicles per minute.

We developed a simulator using MATLAB, and Monte Carlo simulations were performed according to the simulation environments. We ran 10,000 simulations with different random seeds for each case, and averaged the results. Such simulation results, denoted as Simulation in each figure in this section, are compared with analytical results from the driven equations. If the analytical results are obtained from our exact formula, the results are denoted as Analysis. If the analytical results are obtained from our approximation formula, then the results are denoted as Approximation.

A. Connectivity

We first investigate the connectivity of a VANET from various aspects: i) probability distribution of the number of disconnected links, ii) average number of clusters, and iii) probability that the VANET is fully connected for a given number of vehicles. Note that, if we are given information on the number of vehicles at a particular time, i.e., the value of $N(t)$, these three metrics on connectivity are readily characterized by normalized transmission range $a\,(=\frac{R_T}{D})$, as shown in Theorems 1, 2, and 3. In our study, we vary the maximum transmission range of each vehicle as $R_T = 200, 300, 400$ m, so that the corresponding normalized transmission range is $a = 0.067, 0.100, 0.133$.

Fig. 5 shows the probability mass function $\hat{g}_n(\cdot)$ for the number of disconnected links between the RSUs when the number of vehicles is $n = 20$. The graphs denoted as Analysis in Fig. 5 are obtained from (7). From the figure, we can see that the distribution is skewed to the right for each normalized transmission range a. In addition, the peak of each curve moves to the left as a increases, meaning that the larger the normalized transmission range, the more likely it is to have fewer clusters in the VANET.

Fig. 6 details the average number of clusters between the adjacent RSUs when the number of vehicles is $n = 0, 1, \ldots, 70$. The graphs denoted as Analysis in Fig. 6 are obtained from (9). There are two clusters when $n = 0$, since each RSU constitutes a cluster. As the number of vehicles

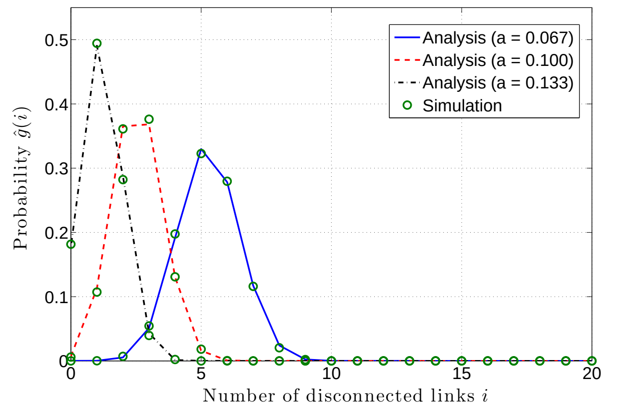


Fig. 5. The probability of having i disconnected links for a network consisting of 20 vehicles when the maximum transmission range is $R_T = 200, 300, 400$ m, and the inter-RSU distance is $D = 3000$ m (i.e., the normalized transmission range is $a = 0.067, 0.100, 0.133$).

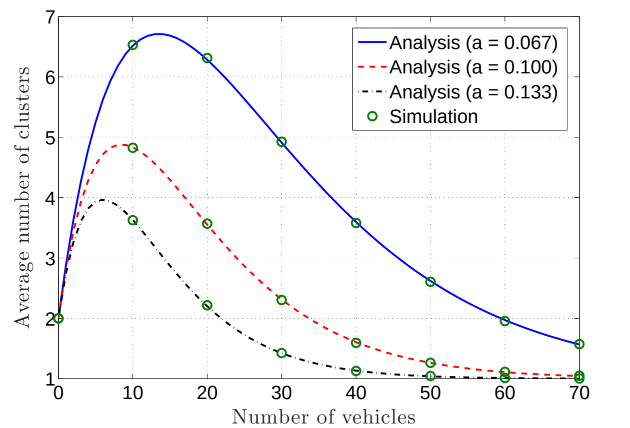


Fig. 6. The expected number of clusters when the maximum transmission range is $R_T = 200, 300, 400$ m and the inter-RSU distance is $D = 3000$ m (i.e., the normalized transmission range is $a = 0.067, 0.100, 0.133$).

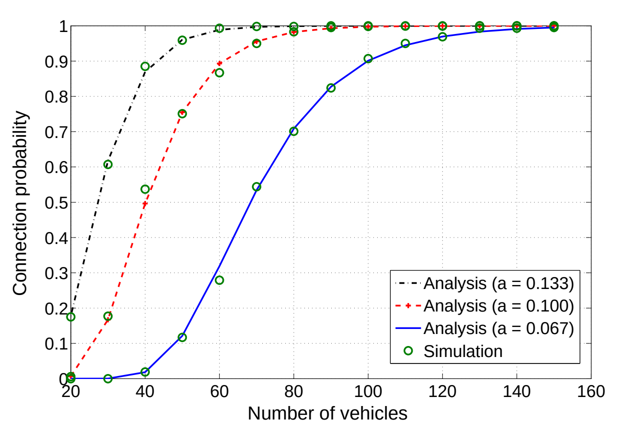


Fig. 7. Connection probability vs. the number of vehicles when the maximum transmission range is $R_T = 200, 300, 400$ m and the inter-RSU distance is $D = 3000$ m (i.e., the normalized transmission range is $a = 0.067, 0.100, 0.133$).

increases, the average number of clusters increases and then decreases, converging to 1 eventually. Hence, the result in Fig. 6 indicates that the number of clusters is always bounded above, and has a maximum. In the integration of a VANET and a cellular network [3], [13], the cluster heads play the role of mobile gateway between the VANET and the cellular network [14]. Hence, the expected number of cluster heads can help to estimate the required capacity and to reserve resources for interworking the VANET with the cellular network. In practice, due to mobility randomness, the number of vehicles between two adjacent RSUs is time-varying. If the number of vehicles is equally likely observed at between 50 and 70 vehicles with a 200 m transmission range, from Fig. 6 and the law of total probability, we can predict that the average number of clusters can be roughly estimated at around two during the observation.

Fig. 7 shows that the connection probability increases as the number of vehicles increases, and saturates at a certain system size. The graphs denoted as Analysis in Fig. 7 are obtained from (11). The connection probability can be interpreted as the time portion of network connection during observation. Hence, combining the results of Figs. 6 and 7 gives us detailed information on the statistical properties of network connectivity as follows. If a VANET is composed of 60 vehicles with a 200 m transmission range (i.e., $a = 0.067$), the network will be fully connected only 31.89% of the time (from Fig. 7) and will be segmented into two clusters on average (from Fig. 6). In addition, network connectivity can be improved as normalized transmission range a increases for each fixed number of vehicles. For example, when the 60 vehicles have a 400 m transmission range (i.e., $a = 0.133$), the probability of network connection increases to 98.86% from Fig. 7, and the average number of clusters becomes 1.01 from Fig. 6. The single cluster of the VANET implies that the network is fully connected most of the time.

Our findings from (9) can apply for multiple RSUs as well, under the assumptions that vehicle arrivals at the first RSU follow a Poisson process and that vehicle velocities chosen randomly at the first RSU are maintained until the last RSU. When $N_{RSU} (\geq 2)$ RSUs are involved, there are $(N_{RSU} - 1)$ segments separated by intervening RSUs. Hence, by separately considering each segment and combining the results, our findings can be extended to such cases. Since $(N_{RSU} - 2)$ clusters including the intervening RSUs are counted twice at two adjacent segments, $(N_{RSU} - 2)$ should be subtracted from the sum of the numbers of clusters individually counted in each segment. For example, if there are l_j disconnections (i.e., $(l_j + 1)$ clusters) in each segment $j \in \{1, 2, \ldots, N_{RSU} - 1\}$ at time t, then the number of clusters in the network, $C(t)$, can be expressed as

$$C(t) = \sum_{j=1}^{N_{RSU}-1} (l_j + 1) - (N_{RSU} - 2) = \sum_{j=1}^{N_{RSU}-1} l_j + 1.$$

B. Average critical network size

We next investigate the average critical network size, $\mathrm{E}[N_a]$, between two adjacent RSUs. Fig. 8 shows the results when the maximum transmission range varies from 100 m to 1500 m (i.e., a varies from 0.033 to 0.5). In the figure, the expectation estimated from our exact formula (14) is denoted as Analysis, and the graph denoted as Approximation shows our asymptotic formula (15) when $o(1) = 0$. As shown in Fig. 8, the average number of vehicles, $\mathrm{E}[N_a]$, required for network connection increases on the order of $\left(\frac{1}{a} \log \frac{1}{a}\right)$ as normalized transmission range a decreases. We also observe that the results obtained by simulation, analysis, and approximation are almost identical. When ratio a is small, the analytical results from (14) deviate from the other results because the combinatorial term of (14) is too large to be included in the simulation, and we consider only a finite number of vehicles in the computation of the infinite series.

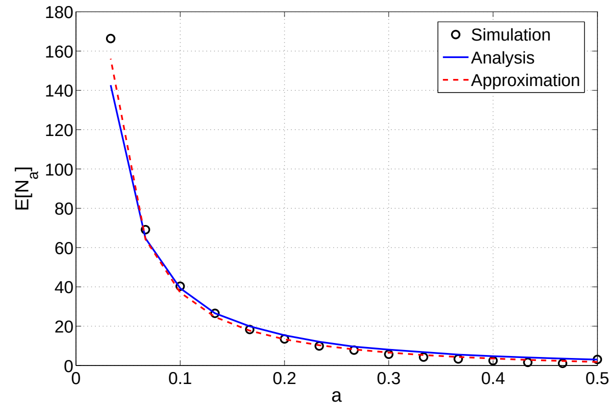


Fig. 8. The expected number of vehicles $\mathrm{E}[N_a]$ vs. normalized transmission range a

As the number of vehicles increases, the network begins a phase transition from a disconnected phase to a connected phase at the critical network size. The critical network size is a function of normalized transmission range a for a given network configuration, and follows $\frac{1}{a}\left(\log \frac{1}{a} + \log \log \frac{1}{a} + 0.5772\right)$ approximately. For example, when $a = 0.1$, the average critical network size is 37. However, if a is halved, i.e., the maximum transmission range is reduced to half of the given condition, or the inter-RSU distance doubles, then the critical network size increases to 94 vehicles, which is greater than the reciprocal of the transmission range reduction.

C. Expected number of active links

To study network capacity, we examine the expected number of simultaneously activated links in a vehicular network as the number of vehicles between the RSUs increases from $n = 20$ to $n = 120$. Fig. 9 shows the results for $a = 0.1, 0.15, 0.2$, where the graph denoted as Approximation is obtained from (25). From the figure, we observe that, as the number of vehicles increases, the number of active links linearly increases with slope $\frac{1}{2}(1 - e^{-2})$, regardless of a. This result indicates that, in an asymptotic regime of n, exactly 43.2% $\left(= \frac{1-e^{-2}}{2}\right)$ of links are simultaneously activated, and the resulting network capacity becomes $0.432n$. In Fig. 9, there is a gap between

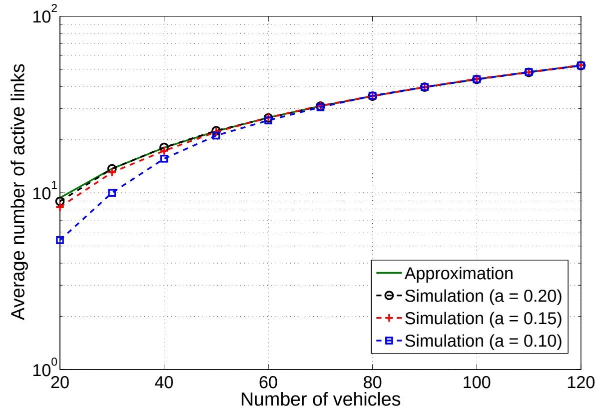


Fig. 9. The expected number of active links vs. the number of vehicles when the normalized transmission range is $a = 0.10, 0.15, 0.20$

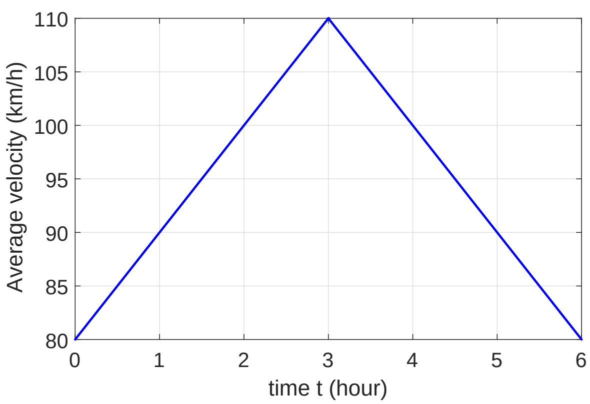


Fig. 10. The average velocity of a vehicle arriving at time t

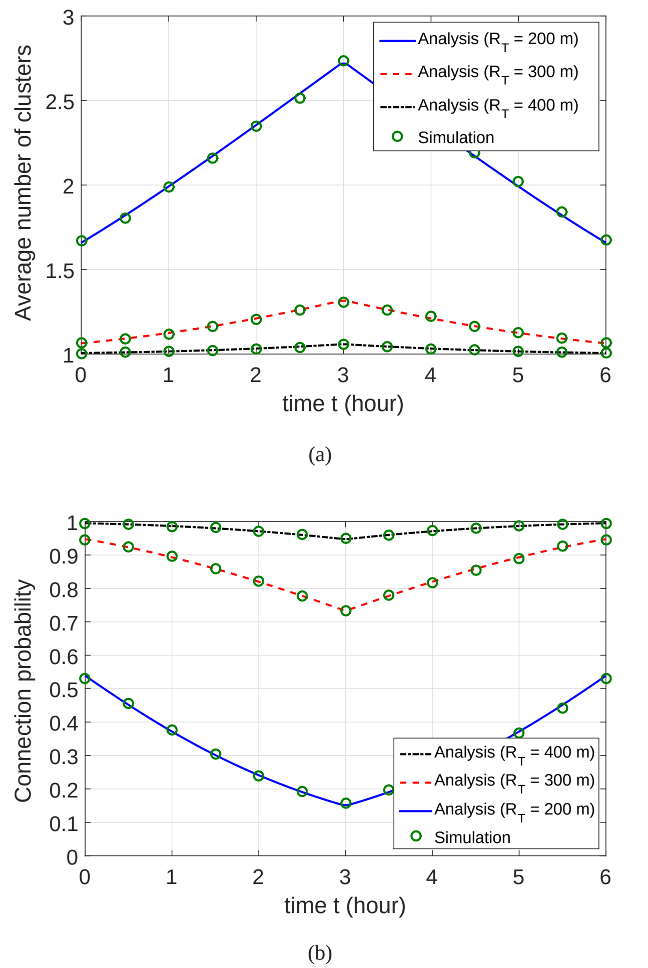


Fig. 11. (a) The expected number of clusters and (b) the connection probability when the average velocity changes over time according to Fig. 10. The maximum transmission range is $R_T = 200, 300, 400$ m, and the inter-RSU distance is $D = 3000$ m.

the results of the simulation and the analysis for a small n. The difference is larger for a smaller a, and becomes zero as n increases. The reason is as follows. Our analysis focuses on the regime $n \gg \mathrm{E}[N_a]$, so that the network is in a fully connected condition, and link activation is ruled only by a random access scheme. However, when the network consists of a small number of vehicles, link connectivity (as well as random access) affects link activation, since a disconnected link cannot be activated. Such an event can occur more often for a smaller a, yielding a larger gap, as shown in Fig. 9.

D. Effect of velocity changes on connectivity

Vehicle mobility pattern is one of the major factors that govern the connectivity of a VANET. In this section, we investigate the effect of average velocity changes on the connectivity of a VANET between two adjacent RSUs.

Let $\mu(t)$ be the average velocity of a vehicle arriving at time t. We assume that $\mu(t)$ changes over time t according to Fig. 10. Each vehicle arriving at time t is assigned a randomly chosen velocity between $(\mu(t) - 20)$ km/h and $(\mu(t) + 20)$ km/h, and maintains its assigned velocity while moving on the highway segment between two adjacent RSUs (i.e., for approximately one to three minutes, since the inter-RSU distance is $D = 3000$ m).[8]

Figs. 11 (a) and (b) show how the average number of clusters, and the connection probability, change when the average velocity varies with time according to Fig. 10. The graphs denoted as Analysis in Figs. 11 (a) and (b) are obtained from (10) and (12), respectively. From the figures, we have the following observations and interpretations: (i) As the average velocity increases, the average number of clusters increases, and the connection probability decreases. This observation indicates that high-speed highways are more prone to disconnection than low-speed highways. Note that, in Section III, we showed that the spatial distribution of vehicles at any given time t is Poisson with the spatial density $\lambda_d = \lambda \mathrm{E}[V^{-1}]$ (see Equation (2)). Hence, for a fixed arrival rate λ, smaller $\mu(t)$ results in a more dense network (having more vehicles), which

[8]In most general cases, the vehicle velocity can vary with time as the vehicle moves along the highway. It has been shown that if vehicle velocities change over time according to a wide-sense stationary random process, the steady-state distribution of vehicle locations is the same as when vehicle velocities are constant over time [11], [38].

in turn leads to improving the connectivity, as shown in Figs. 6 and 7. (ii) The amount of increase or decrease in Figs. 11 (a) and (b) becomes prominent for shorter transmission range R_T. That is, the shorter the transmission range, the connectivity becomes more sensitive to velocity changes. Hence, using a longer transmission range can help to stably sustain network connectivity over time against velocity changes. (iii) In the worst case (i.e., $R_T = 200$ m), the connection probability is in the range from 0.15 to 0.55 with 1.7 to 2.7 clusters on average. The connectivity can be noticeably improved by increasing the transmission range from $R_T = 200$ m to $R_T = 300$ m; the network is in a fully connected state for at least 4 out of 6 hours with 1 to 1.3 clusters on average.

VII. Conclusion and Future Work

In this paper, we analyze properties of a vehicle-to-vehicle network with a random topology. To that end, we model a one-dimensional random topology network with geometric elements, such as line segments and points. Using theory from geometric probability, we analyze properties of such a vehicular network not only in a disconnected phase but also in a connected phase. For a given number of vehicles, our analysis provides the probability that the network is connected, as well as the expected number of clusters that the network forms. Moreover, our analysis shows that the average number of vehicles needed for the network to be fully connected is approximately $\frac{1}{a}\left(\log\frac{1}{a} + \log\log\frac{1}{a} + 0.5772\right)$ for a given transmission range, R_T, where $a = \frac{R_T}{D}$. When CSMA/CA is adopted as a random access scheme, then each packet between RSUs has $\frac{2e^2}{e^2-1}n$ slot times. We verified our analysis through simulations.

Our findings focused on a VANET between two adjacent RSUs on a unidirectional highway under the assumption that the maximum transmission range is deterministic. However, due to wireless channel randomness, such as shadowing and fading, the transmission range has randomness, which affects the properties of network connectivity. Interference among transmissions and unevenly distributed-traffic congestion may affect the network connectivity as well. Bidirectional traffic induces a different network topology than unidirectional traffic. The analysis of a VANET between multiple RSUs with time-varying vehicle velocities requires other mathematical approaches, since the location of a vehicle is no longer linear with time. The analyses extending to the above assumptions are left open for future work.

References

[1] H. Hartenstein and K. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Communications Magazine*, vol. 46, no. 6, pp. 164–171, 2008.

[2] F. Hoque and S. Kwon, "An emergency packet forwarding scheme for V2V communication networks," *The Scientific World Journal*, vol. 2014, pp. 1–7, June 2014.

[3] A. Benslimane, T. Taleb, and R. Sivaraj, "Dynamic clustering-based adaptive mobile gateway management in integrated VANET–3G heterogeneous wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 3, pp. 559–570, 2011.

[4] B. Hassanabadi, C. Shea, L. Zhang, and S. Valaee, "Clustering in vehicular ad hoc networks using affinity propagation," *Ad Hoc Networks*, vol. 13, pp. 535–548, 2014.

[5] IEEE, "IEEE Standard 802.11p: Wireless lan medium access control and physcal layer specification," 2012.

[6] P. Santi and D. M. Blough, "The critical transmitting range for connectivity in sparse wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 1, pp. 25–39, 2003.

[7] N. Wisitpongphan, F. Bai, P. Mudalige, V. Sadekar, and O. Tonguz, "Routing in sparse vehicular ad hoc wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 8, pp. 1538–1556, 2007.

[8] M. Khabazian and M. K. M. Ali, "A performance modeling of connectivity in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 4, pp. 2440–2450, 2008.

[9] S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy, "Improving connectivity in vehicular ad hoc networks: An analytical study," *Computer Communications*, vol. 31, no. 9, pp. 1653–1659, 2008.

[10] J. Wu, "Connectivity of mobile linear networks with dynamic node population and delay constraint," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1218–1225, 2009.

[11] S. Durrani, X. Zhou, and A. Chandra, "Effect of vehicle moblity on connectivity of vehicular ad hoc networks," in *Proceedings of IEEE VTC 2010*, 2010.

[12] N. P. Chandrasekharamenon and B. AnchareV, "Connectivity analysis of one-dimensional vehicular ad hoc networks in fading channels," *EURASIP Journal on Wireless Communications and Networking*, no. 1, pp. 1–16, 2012.

[13] G. Araniti, C. Campolo, M. Condoluci, A. Iera, and A. Molinaro, "LTE for vehicular networking: A survey," *IEEE Communications Magazine*, vol. 51, no. 5, pp. 148–157, 2013.

[14] G. Rémy, S.-M. Senouci, F. Jan, and Y. Gourhant, "LTE4V2X-collection, dissemination and multi-hop forwarding," in *Proceedings of IEEE ICC'12*, 2012.

[15] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic engineering*. Pearson Prentice Hall, 2004.

[16] B. Bollobas and O. Riordan, *Percolation*. Cambridge, UK: Cambridge University Press, 2009.

[17] G. Grimmett, *Probability on graphs: random processes on graphs and lattices*. Cambridge University Press, 2010, vol. 1.

[18] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.

[19] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, 2002.

[20] S. N. Diggavia, M. Grossglauser, and D. Tse, "Even one-dimensional mobility increases the capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 51, no. 11, pp. 3947–3954, 2005.

[21] G. H. Mohimani, F. Ashtiani, A. Javanmard, and M. Hamdi, "Mobility modeling, spatial traffic distribution, and probability of connectivity for sparse and dense vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1998–2007, 2009.

[22] S. Yousefi, E. Altman, R. El-Azouzi, and M. Fathy, "Analytical model for connectivity in vehicular ad hoc networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 6, pp. 3341–3356, 2008.

[23] J. Gomez and A. T. Campbell, "A case for variable-range transmission power control in wireless multihop networks," in *Proceedings of IEEE INFOCOM 2004*, 2004.

[24] S. Kwon and N. B. Shroff, "Energy-efficient unified routing algorithm for multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3890–3899, 2012.

[25] A. Nasipuri and S. R. Das, "Multichannel CSMA with signal power-based channel selection for multihop wireless networks," in *Proceedings of IEEE VTC'00 Spring*, 2000.

[26] J. So and N. H. Vaidya, "Multi-channel MAC for ad hoc networks: handling multi-channel hidden terminals using a single transceiver," in *Proceedings of ACM MobiHoc'04*, 2004.

[27] H. F. Q. Wang, S. Leng and Y. Zhang, "An IEEE 802.11p-based multichannel MAC scheme with channel coordination for vehicular ad hoc networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 449–458, 2012.

[28] C. Ryll-Nardzewski, "Remarks on the Poisson stochastic process (III) (On a property of the homogeneous Poisson process)," *Studia Mathematica*, vol. 14, no. 2, pp. 314–318, 1954.

[29] L. Breiman, "The Poisson tendency in traffic distribution," *Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 308–311, 1963.

[30] S. M. Ross, *Stochastic process*, 2nd ed. John Wiley and Sons, 1996.

[31] W. Stevens, "Solution to a geometrical problem in probability," *Annals of Eugenics*, vol. 9, no. 4, pp. 315–320, 1939.

[32] A. F. Siegel and L. Holst, "Covering the circle with random arcs of random sizes," 1981, technical report.

[33] G. B. Kunin, "The finite difference calculus and applications to the interpolation of sequences." [Online]. Available: http://www-math.mit.edu/phase2/UJM/vol1/GKUNIN 1.PDF

[34] H. Solomon, *Geometric probability*. Philadelphia: SIAM, 1978.

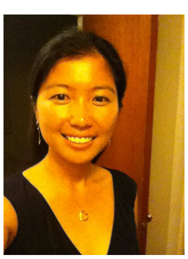
[35] E. W. Weisstein, "Euler-mascheroni constant." [Online]. Available: http://mathworld.wolfram.com/Euler-MascheroniConstant.html

[36] F. W. Steutel, "Random division of an interval," *Statistica Neerlandica*, vol. 21, no. 3-4, pp. 231–244, 1967.

[37] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 13, pp. 535–547, 2000.

[38] J. Wu, "Connectivity analysis of a mobile vehicular ad hoc network with dynamic node population," in *Proceedings of IEEE Globecom Workshops'08*, 2008.

Sungoh Kwon received the B.S. and M.S. degrees in electrical engineering from KAIST, Daejeon, Korea, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, in 1994, 1996, and 2007, respectively. From 1996 to 2001, he was a research staff member with Shinsegi Telecomm Inc., Seoul, Korea. From 2007 to 2010, he developed LTE schedulers as a principal engineer in Samsung Electronics Company, Ltd., Korea. He has joined to University of Ulsan as an Assistant Professor since 2010 and is now an Associate Professor. His research interests are in wireless communication networks.



Yoora Kim received her B.S., M.S. and Ph.D. degrees in Mechanical Engineering, Applied Mathematics and Mathematical Sciences from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 2003, 2005 and 2009, respectively. She is currently an Assistant Professor with the Department of Mathematics at University of Ulsan, Ulsan, South Korea. Prior to joining University of Ulsan, she has been with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, as a Post-Doctoral Research Associate. Her research interests include modeling, design, and performance evaluation of communication systems and scheduling and resource allocation problems in wireless networks under various stochastic dynamics.

Ness B. Shroff received his Ph.D. degree from Columbia University, NY in 1994 and joined Purdue university immediately thereafter as an Assistant Professor. At Purdue, he became Professor of the school of Electrical and Computer Engineering in 2003 and director of CWSA in 2004, a university-wide center on wireless systems and applications. In July 2007, he joined the Ohio State University where he holds the Ohio Eminent Scholar Chaired Professorship in Networking and Communications. His research interests span the areas of communication, networking, storage, cloud, recommender, social, and cyberphysical systems. He is especially interested in fundamental problems in learning, design, control, performance, pricing, and security of these complex systems. Dr. Shroff currently serves as editor-at-large in the IEEE/ACM Trans. on Networking, and as senior editor of the IEEE Transactions on Control of Networked Systems. He also serves on the editorial board of the IEEE Network Magazine. He has served on the technical and executive committees of several major conferences and workshops. He was the TPC co-chair of IEEE INFOCOM03, ACM Mobihoc08, and general chair of IEEE CCW99 and WICON08. Dr. Shroff is a Fellow of the IEEE. He has received numerous awards for his work, including two best paper awards at IEEE INFOCOM (in 2006 and 2008), the flagship conference of the field. He has also received the IEEE WiOPT 2013, IEEE WiOpt 2012, and the IWQoS best student paper award, the 2005 best paper of the year award for the Journal of Communications and Networking, the 2003 best paper of the year award for Computer Networks, and the NSF CAREER award in 1996 (his IEEE INFOCOM 2013 and IEEE INFOCOM 2005 papers ware selected as runner-up papers). Dr. Shroff is among the list of highly cited researchers from Thomson Reuters ISI and in Thomson Reuters Book on The Worlds Most Influential Scientific Minds. In 2014, he received the IEEE INFOCOM achievement award for seminal contributions to scheduling and resource allocation in wireless networks.