Capturing the Spread of Information in Heterogeneous V2X through Scalable Computation

Jungyeol Kim, Rohan Saraogi, Saswati Sarkar, David Starobinski, and Santosh S. Venkatesh

***Abstract*—Emerging V2X technology enables vehicles to exchange messages with each other (V2V) and with signaling infrastructure (I2V) on the roadways. Information propagation in transportation networks is highly influenced by both vehicle mobility and wireless communication. As for vehicle mobility, realistic traffic flow changes with time, exhibiting sharp time-triggered transitions, due to external factors such as traffic lights. Thus, mobility process is temporally heterogeneous and not smooth, which fundamentally alters the dynamics of V2X (V2V and I2V together) message propagation in a complex manner. As for wireless communication, communication heterogeneity is an integral component of V2X systems - different types of vehicles may have different communication capabilities, and V2V and I2V communications coexist. We propose a mathematical framework, based on a continuous-time Markov chain (CTMC), for characterizing the spatio-temporal spread of V2X information (1) when the traffic flow exhibits sharp time-triggered transitions and (2) when there exists communication heterogeneity comprising of different V2V commutation capabilities, different wireless communication conditions, and both V2V and I2V. We prove that the state evolutions under the CTMC model converge to a set of differential equations in the asymptotic limit of a large number of vehicles, enabling computations that gracefully scale with increase in network size and the number of vehicles. Our framework can accommodate arbitrary traffic synchronization patterns corresponding for example to incorporate the presence of an arbitrary number of traffic signals. Furthermore, numerical computations using this mathematical framework answer several questions that influence the practice of V2X network design and security.**

I. Introduction

Vehicles can share information with each other via vehicle-to-vehicle communication (V2V) or with infrastructures on the roadway via infrastructure-to-vehicle communication (I2V). These are collectively referred to as V2X. We seek a mathematical framework that characterizes the fraction of vehicles that have received data through V2X, equivalently the probability distributions, as a function of space and time, in an arbitrary road topology. In vehicular networks, the evolution of mobility and communication is intrinsically stochastic, which leads to substantial challenges in computing the relevant probability distributions. Computation approaches are available only for steady-state distributions of even regular stochastic processes such as Markov processes, but even these approaches are computationally intensive because they rely on the inversion of the transition probability matrices with dimensions increasing with the number of vehicles and size of the road network. We seek to obtain the distribution of the number of vehicles that carry the information at any finite time, which is much more challenging than computing the steady-state distributions. We recently characterized the same for the simple scenario of homogeneous mobility and communication: specifically 1) the mobility process does not depend on time and 2) all vehicles are homogeneous concerning communication and there is no I2V. [1].

However, in real systems, the mobility process changes with time not only through traffic congestion but also through sharp time-triggered transitions due to external factors such as traffic lights, unpredictable disruptions (e.g., accidents), or planned disruptions (e.g., road-block). For physical traffic lights, all vehicles stop when the light turns red, and they start moving when the light turns green. For virtual traffic lights, the simultaneous stopping and the resumption of movement are both accomplished through the exchange of V2V messages amongst the vehicles (refer to [2] for an example of a virtual traffic light protocol). Thus the mobility process is temporally *heterogeneous* and not smooth. Both physical and virtual traffic lights alter the dynamics of V2V message flow because of the following reasons: 1) all the vehicles simultaneously stop during a certain period and subsequently move during the next period, thus traffic flow becomes synchronized, and 2) different kinds of V2V messages can be quickly propagated amongst vehicles that wait in close proximity to each other at the red traffic light (or virtual equivalent). Ignoring these effects in the model lead to a significant divergence between the message propagation pattern obtained from the model and what arises in practice (e.g., see Figure 1). Note that Figure 1 show that the divergence can be significant even with one traffic light. The divergence may increase significantly more when there are multiple traffic lights due to the correlation between red and green traffic lights. Transportation networks in large cities have a large number of traffic lights (e.g., 2,820 intersections in Manhattan alone are controlled by traffic signals as of June 2011 [3]).

Furthermore, as for communication, *heterogeneity* constitutes the very essence of V2X systems – different vehicles may have different communication capabilities, experience different communication ambience and execute different communica-

This research was supported in part by NSF under grants CCF-2008284 and CCF-2006628.

Saswati Sarkar and Santosh S. Venkatesh are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (email: swati@seas.upenn.edu; venkates@seas.upenn.edu). Jungyeol Kim was with the Department of Electrical and Systems Engineering, University of Pennsylvania (email: jungyeol@alumni.upenn.edu). He is now with JPMorgan Chase. Rohan Saraogi is with School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104 USA (email: rohan.saraogi@gmail.com). David Starobinski is with the Division of Systems Engineering, Boston University, Boston, MA, 02215, USA (email: staro@bu.edu).

tion strategies. For example, either due to intrinsic hardware or software limitations, or strategic choice towards enhancing resilience to cyber-attacks, only some vehicles may be allowed to both transmit and receive information, while the remaining vehicles may only receive information. Note that cyber-attacks on V2X systems can abuse anonymous authentication techniques [4], cause significant delays, or send manipulated messages to attack the vehicular network. Resilience may be enhanced by allowing only some authenticated and secured vehicles to transmit, particularly during the dispersal of information of high security-value such as certificate revocation lists (CRLs)[1]. Also, if only a few vehicles transmit, then when a cyber-attack is suspected diagnostics need to be run only on a few vehicles and the malefactors can be detected in a short time [8]. In view of all the above, we allow for scenarios in which only a certain proportion of vehicles are allowed to both forward and receive the information (e.g., CRL) via V2V with aid of RSUs while the rest can only receive the information. The first set is referred to as *certified*, to indicate that they have been verified or pronounced to be more trustworthy and have the requisite capability to digitally sign; the second set of vehicles is referred to as *non-certified*. The central certification authority determines which vehicles are certified and maintains a list thereof.

Infrastructure is another essential element of the V2X system. I2V communications from road side units (RSUs) bolsters the spread of messages beyond V2V communication, and also supports vehicular network applications such as smart traffic lights, road signing, emergency broadcasts, road traffic advisory. Most infrastructure units do not disclose internal information without the approval of the central authority, while on-board units installed in each vehicle can be easily attacked by malicious agents. Thus, I2V is inherently more secure than V2V. But, deploying RSUs, or converting existing road side infrastructure to RSUs, incurs additional expense, and renders V2X communication even more heterogeneous as V2V fundamentally differs from I2V.

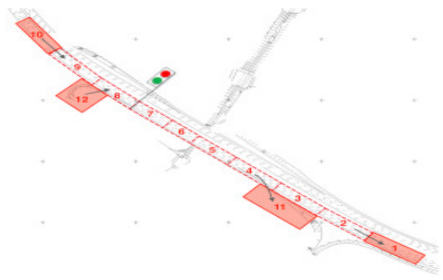
As for the temporally heterogeneous mobility process, we seek to characterize the impact of traffic signals on the fraction of vehicles that have received a message as a function of time and space. As for heterogeneity in V2X communications, the following questions arise: (1) How do the different vehicles with different communication capabilities affect the spread of V2X messages? What percentage of vehicles capable of transmitting would ensure that the desired propagation speed is attained? (2) How do diverse communication conditions, such as interference due to the high density of transmitting vehicles, affect the spread of V2X messages? (3) How does infrastructure affect information propagation? Which is better in terms of propagation speed - greater number of the infrastructure units or vehicles capable of transmitting? The answers to these questions, particularly (1) and (3) would depend on the prevailing communication and mobility conditions which change with time and location. The answers would also help determine the appropriate trade-offs between expense, security risks and V2X message propagation efficacy. Developing a mathematical framework for tractable computation of the distribution of informed vehicles as a function of space and time in presence of mobility and communication heterogeneities is a prerequisite for answering these questions. This remains an uncharted terrain, which we seek to contribute to in a mathematically rigorous albeit computationally tractable manner.

We now describe the related work. A genre of work designs traffic signal control to improve traffic flow and energy consumption at intersections. For example, [9] provides an intelligent traffic management system in two types of road intersections (roundabouts and crossroads) supported by V2X. [10], [11] demonstrate that energy consumption at signalized intersections can be optimized by reducing vehicle idling through communication between infrastructures and vehicles. Also, [12], [13] apply deep reinforcement learning to traffic light control problems based on the Markov decision process framework. However, this genre does not consider the flow of V2X messages. Another genre of work studies propagation speed in presence of communication heterogeneity. [14], [15] considers V2V-equipped and unequipped vehicles, and studies the speed of information spread amongst equipped vehicles at each time and location using traffic flow that is present at that time and nearby. In the study, equipped vehicles are related to both information dissemination and traffic flow, while unequipped vehicles are only directly tied to traffic flow dynamics. However, to the best of our knowledge, this genre does not capture the fraction of vehicles that have received V2X messages as a function of space and time in an arbitrary road topology. The third genre of analytical studies have focused on analyzing propagation speed along one-dimensional road in V2V system for given vehicle speed, traffic density, and distance between vehicles [16]–[21]. This genre has seen limited research on two-dimensional road systems. A recent paper [22] studies the delay in message forwarding along a selected path in a two-dimensional road system under the assumption that all vehicles traveling in the same direction on a road segment are traveling at the same speed. In this paper, the authors introduced an algorithm to choose the path with the minimum expected delay. Another study [23] approximates the speed of information dissemination "wave" at each time

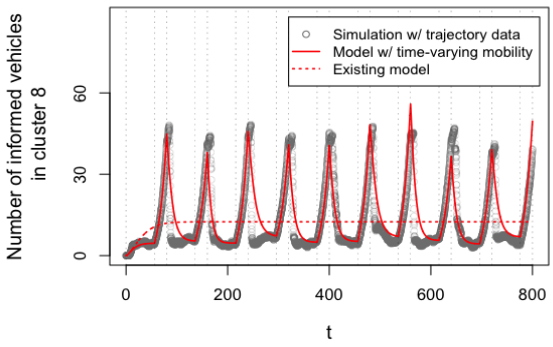
[1]There are simulation studies that address the efficient distribution of information through V2X in vehicular networks, with a particular focus on the distribution of Certificate Revocation List (CRL). A certification authority (CA) is responsible for incorporating the identification of the invalidated certificate(s) into a CRL. Subsequently, the CRL update can be initiated by Roadside Units (RSUs), whereby they disseminate the CRL to vehicles that pass by. A study [5] has proposed the utilization of V2V communication as a strategy to enhance the efficient dissemination of CRLs. This approach enables vehicles that have received the CRL updates to transmit it to other vehicles they encounter, thereby disseminating the information more rapidly throughout the network. A study [6] reestablishes the results in [5], but for partial deployment scenarios of V2V communication, which refer to situations where only a specific proportion of vehicles are equipped with VANET radios. Note that dissemination of CRLs through V2V communication introduces certain security risks, e.g., if a vehicle that forwards the CRL alters it before forwarding the security of the entire system would be compromised. Thus vehicles that forward the CRL should have a certain level of security clearance which only a small fraction of the vehicles may possess as the forwardings by these vehicles need to be periodically verified to ensure that they remain trustworthy. Also, if a vehicle forwards the CRL it receives, it ought to digitally sign the CRL, a capability that not all vehicles possess in general. Finally, if an extensive volume of information distributions takes place and is communicated throughout the entire network, it has the potential to induce network congestion or impede the provision of other essential V2X messages with higher priority [7].

and location using only the deterministic traffic flow that is present at that moment and nearby. As such, this genre of work has mostly concentrated on estimating propagation speed (or expected delivery delays), that too without any facilitating infrastructure and for temporally smooth mobility processes, which can happen only when there are no traffic signals. We seek characterization of spatio-temporal distribution of messages in V2X systems, i.e., fraction of vehicles who have received the desired message, as a function of space and time, over two-dimensional roads with infrastructure and traffic signals. This has hitherto remained an unchartered territory. The closest to the current work is our previous work [1] which characterizes the spatio-temporal distribution of messages in V2X systems, but under the limitation that there is no infrastructure and the mobility process is temporally smooth. The generalization in the above dimensions, which is critical for capturing practical attributes of transportation systems, calls for fundamental methodological innovations, which we attain in this paper.

We consider pulsed vehicular traffic flow due to traffic signals, which leads to mobility heterogeneity as discussed above. We also consider a V2X network with (1) two classes of vehicles differentiated by their communication capabilities: *certified* vehicles capable of both transmitting and receiving, and *non-certified* vehicles that can only receive, (2) infrastructures in the form of RSUs at various locations. We start from a continuous-time Markov chain (CTMC) and present a mathematical model for characterizing the spatio-temporal spread of V2X information under the aforementioned conditions. As for the temporally heterogeneous mobility process, we capture the time-varying mobility by combining continuous evolution of state variables with discrete, instantaneous changes of mobility rates between red and green traffic lights. This constitutes the first methodological innovation. Despite the fact that the mobility functions of the equations are instantaneously changed, executions of the system are always continuous. The second methodological innovation is to accommodate infrastructure in the model despite fundamental differences between V2V and I2V communications. We also capture the different classes of vehicles with different communication capabilities by introducing different state variables where each variable represents each class. We then show that, despite the presence of mobility and communication heterogeneities, as the number of vehicles approaches infinity, the stochastic model converges to a system of differential equations, which provides the fraction of vehicles that have received an information of interest, as a function of time and space (Section II). The computation time scales efficiently with the number of overall vehicles and the size of the transportation network and the number of RSUs, which constitutes a significant strength of the approach given the scale of modern transportation networks. In Section III, we use the differential equation formulation to answer specific questions that heterogeneity in mobility and communication pose. In Section IV, we articulate the bearings of our findings on the practice of V2X network design and security and present a direction of future research to address a limitation.



(a)

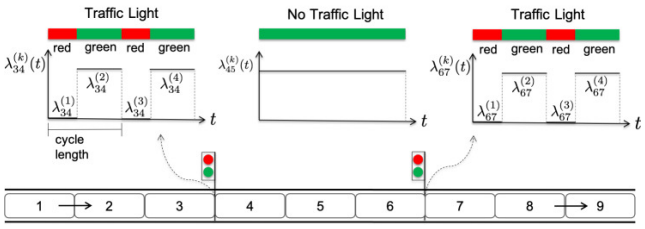


(b)

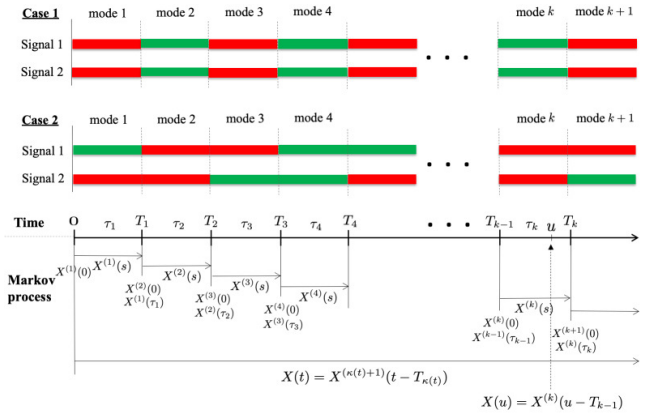
Fig. 1. (a) Clustered U.S Highway 101 with single hypothetical traffic light. We assume that there is a traffic light between clusters 7 and 8. (This figure is modified from Figure 11a of [1]) (b) Fraction of informed vehicles in cluster 8 over time. The gray points are the average of 100 simulation runs for V2V message flow. The message flow has been obtained by superimposing a statistical communication process on a synthetic trace data, modified based on an actual trajectory dataset [24], to reflect the pulsed traffic. The dotted red line is the solution of the previous model in [1], which does not incorporate traffic lights. Notice that there is a significant difference between the gray points and the dotted red line. The solid red line is the solution of the model that we will present in this paper. The solid red line tracks the gray points more closely than the the dotted red line. The ratio between the average deviation over time between the gray points and the dotted red line and that between the gray points and the solid red line is approximately 2.5:1.

II. Model Formulation

We propose a model with a macro-level view of the vehicular networks where the system is expressed for aggregates rather than for individual vehicles. The overall road topology is divided into J clusters, where each cluster corresponds to a specific part of the road (Figure 3 for an example). Our model groups vehicles in clusters, allows them to move across clusters, and communicate within and across clusters. The model can accommodate vehicular mobility conditions including pulsed traffic and V2X (V2V and I2V) communication conditions. We show that the stochastic model converges to a set of differential equations (diffusion equations) as the number of vehicles increases, despite discontinuous temporal changes in the traffic flow and communication heterogeneity. Our framework scales efficiently to a large number of vehicles in a large-scale vehicular network (i.e., metropolitan city) while most vehicular network simulators (e.g., VEINS[25]) can only realistically simulate small to medium number of vehicles due to large memory usage. We consider the prop-



(a)



(b)

Fig. 2. (a) (This figure is modified from Figure 14 of [1]) Clustered road topologies. The mobility rates between clusters with traffic lights are changed discontinuously depending on the mode switching over time. (b) Evolution of the process. Case 1 is an example of two traffic lights being synchronized, and case 2 is an example of not being synchronized.

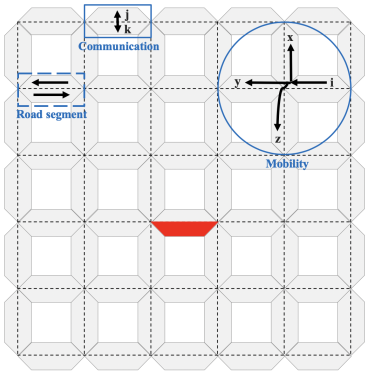


Fig. 3. Clustered grid road topology with bidirectional traffic. Each road segment consists of 2 clusters corresponding to the opposite directional roads (see blue dashed rectangle). The mobility and communication networks can be described through routing probabilities and communication rates between pairs of clusters. The blue solid circle represents a part of mobility network, which elucidates clusters x, y, z to which vehicles can move from an example cluster i; here, routing probability $p_{il}(\cdot) = 1/3$ for $l \in \{x, y, z\}$, and $p_{il}(\cdot) = 0$ if $l \notin \{x, y, z\}$. The blue solid rectangle represents a part of communication network. It is an example pair of clusters, j, k on the same road segment, and vehicles in them can directly communicate to each other through V2V. Thus, communication rate $\beta_{jk}(\cdot) > 0$, and $\beta_{jl}(\cdot) = 0$ all l such that $l \notin \{j, k\}$. Note that the mobility and communication networks overlap but may not be identical. Vehicles in the two adjoining clusters j, k are able to directly communicate between each other ($\beta_{jk}(\cdot) > 0$); but physical travel between the clusters is not possible ($p_{jk}(\cdot) = p_{kj}(\cdot) = 0$) because traffic regulations prohibit vehicles from crossing the median. (This figure is adapted from [1])

agation of a message of interest through V2X. The message of interest can for example be traffic conditions, road safety, and CRLs (more examples of messages of interest can be found in [1]). From hence, we call vehicles that have received the V2X message of interest as 'informed vehicles', while referring to vehicles that have not received such messages as 'non-informed vehicles'.

A. Heterogeneity in mobility process - Sharp time-triggered transitions

We capture time-varying mobility by combining continuous evolution of state variables with discrete, instantaneous changes. The time axis is now subdivided into a sequence of *modes*, such that the mobility rates switch discontinuously to different values at the end of each mode, and remain constant during each mode. For example, the mode can be the time interval between color changes of traffic lights. We illustrate the concept of mode through the following example.

Example 1. *Consider a transportation network with two traffic lights (Figure 2a). Whenever any one traffic light changes color, the mode changes (Figure 2b). If the two traffic lights are synchronized, the mode changes when the signals of both traffic lights are changed at the same time (Case 1 in Figure 2b). In this case, the mobility rates discontinuously transition when both signals change simultaneously; the mobility rates in the clusters with traffic lights are 0 if the lights are red, and other specified values if they are green. Figure 2a provides an example of transitions in mobility rates with changes in traffic lights. If the two traffic lights are not synchronized, the mode changes when either signal changes, and the mobility rate in the cluster with the corresponding traffic light transitions from* 0 *to the normal value or vice versa (Case 2 in Figure 2b).*

Let $\mathbf{n}^{I:k}(s)$ and $\mathbf{n}^{S:k}(s)$ represent J-dimensional vectors, whose j-th elements are the number of informed and non-informed vehicles respectively, in cluster j, at time s from the beginning of the kth mode, $k \in \mathbb{Z}_{>0}$. Consider the state process $X^{(k)}(s) = (\mathbf{n}^{I:k}(s), \mathbf{n}^{S:k}(s))$. The time duration of the k-th mode, τ_k, is defined as the time elapsed between the $(k-1)$-th and k-th signal changes, and T_k denotes the time at which the signal changes for the k-th time, thus τ_k and T_k are given by

$$\tau_k = T_k - T_{k-1}; \quad T_k = \textstyle\sum_{j=1}^{k} \tau_j, \quad T_0 = 0.$$

The last state of the k-th mode is considered the initial state of the next mode $k+1$ (i.e., $X^{(k+1)}(0) = X^{(k)}(\tau_k)$). This process is illustrated in detail in Figure 2b.

We now consider the $2J$-dimensional vector $X(t) = (\mathbf{n}^I(t), \mathbf{n}^S(t))$. The state process $X(t) = \left(n_1^I(t), n_2^I(t), \ldots, n_J^I(t);\ n_1^S(t), n_2^S(t), \ldots, n_J^S(t)\right)$ represents the state at time t from the beginning of the whole process, not just from the beginning of each mode; the first J elements represent the number of informed vehicles for each cluster, and the remaining J elements after the semicolon represent the number of non-informed vehicles for each cluster. Naturally, the snapshot of the process at time t from the beginning of the whole process, $X(t)$, is given by

$$X(t) = X^{\kappa(t)+1}\left(t - T_{\kappa(t)}\right),$$

where $\kappa(t) = \max\{k|T_k < t\}$ is the total number of signal changes in the system by time t, regardless of which traffic light it is. The state space is the set of lattice points in $\mathbb{Z}^J \times \mathbb{Z}^J$ satisfying

$$S^N := \left\{(\mathbf{n}^I, \mathbf{n}^S) \mid n_j^I \geq 0,\ n_j^S \geq 0,\ j = 1, \ldots, J;\ \textstyle\sum_{j=1}^J \left(n_j^I + n_j^S\right) = N\right\}.$$

In each mode, we assume that the time until a vehicle in a cluster moves to a neighboring cluster is exponentially distributed with parameters depending on the states. Similarly, for both intra- and inter-cluster communication, we assume that the time between successful communications from an informed vehicle to a non-informed vehicle is exponentially distributed. Under these assumptions, the process $X^{(k)}(t)$, $k \in \mathbb{Z}_{>0}$, becomes a continuous-time Markov chain (CTMC).

The CTMC has the following three state transitions: 1) an informed vehicle moves from cluster i to cluster j, $j \neq i$; 2) a non-informed vehicle moves from cluster i to cluster j, $j \neq i$; and 3) a non-informed vehicle in a cluster j receives an information from an informed vehicle located in the same cluster or in a different cluster.

The first two types of state transitions capture the mobility of the vehicle in the system, which can vary depending on the mode. During the k-th mode, an informed and non-informed vehicle migrate along the road departing from the cluster i to j at rate $\lambda_{ij}^{I:k}(\cdot)$ and $\lambda_{ij}^{S:k}(\cdot)$, respectively. We assume that for every mode $k \in \mathbb{Z}_{>0}$, both $\lambda_{ij}^{I:k}(\cdot)$ and $\lambda_{ij}^{S:k}(\cdot)$ are bounded functions of $\frac{1}{N}(\mathbf{n}^{I:k}, \mathbf{n}^{S:k})$ if vehicle movement is possible from cluster i to j and 0 otherwise. Specifically, if the traffic light is located between clusters i and j, the values of the parameters $\lambda_{ij}^{I:k}(\cdot)$ and $\lambda_{ij}^{S:k}(\cdot)$ change along with the lights: they are 0 if the lights are red in mode k, and at the normal values if they are green in mode k (Figure 2a). Finally, the last type of state transition captures the successful communication between an informed vehicle and a non-informed vehicle. A vehicle in cluster i successfully communicates with a vehicle in the same cluster i at rate $\beta_{ii}^{(N)}$. A vehicle in cluster i can also successfully communicate with a vehicle in a distinct cluster j at rate $\beta_{ij}^{(N)}$ if the distance between clusters i and j is within communication range and 0 otherwise. The transitions, therefore, capture distinctions in the vehicular routing choices, vehicular speeds, their communication choices etc. based on whether vehicles have the message or not and vehicular congestion in local clusters.

The CTMC can be approximated by a set of ordinary differential equations in the continuum. We define a set $E := \left\{(\mathbf{I}, \mathbf{S}) \mid I_i \geq 0, S_i \geq 0, i = 1, 2, \ldots, J;\ \sum_{i=1}^J (I_i + S_i) = 1\right\}$ with $\left(\mathbf{I}, \mathbf{S}\right) = (I_1, I_2, \ldots, I_J;\ S_1, S_2, \ldots, S_J)$. Define $\beta_{ij} := N\beta_{ij}^{(N)}$, and suppose β_{ij} is constant. There is an implicit assumption that the larger the total number of vehicles, the lower is the communication rate $\beta_{ij}^{(N)}$ which holds when the bandwidth is limited.

We now introduce the formal notation $\lim_{N\to\infty} \frac{\mathbf{n}^I(t)}{N} = \mathbf{I}(t)$ and $\lim_{N\to\infty} \frac{\mathbf{n}^S(t)}{N} = \mathbf{S}(t)$. Note that $\mathbf{I}(t)$ and $\mathbf{S}(t)$ respectively represent the fraction of informed and non-informed vehicles in each cluster. The following theorem provides sufficient conditions for the convergence of the scaled process $X_N(t) = \frac{1}{N}\left(\mathbf{n}^I(t), \mathbf{n}^S(t)\right)$ to a solution of differential equations, $\mathbf{x}(t) = \left(\mathbf{I}(t), \mathbf{S}(t)\right)$, despite the mobility rates changing discontinuously depending on the mode.

Theorem 1. *Suppose for $i, j = 1, 2, ..., J$ and $i \neq j$, mobility rate functions $\lambda_{ij}^{I:k} : E \to \mathbb{R}$ and $\lambda_{ij}^{S:k} : E \to \mathbb{R}$ in every mode $k \in \mathbb{Z}_{>0}$ are bounded and Lipschitz continuous on E. Let $\left(\mathbf{I}(0), \mathbf{S}(0)\right) = \lim_{N\to\infty} \frac{1}{N}\left(\mathbf{n}^I(0), \mathbf{n}^S(0)\right)$, and $\left(\mathbf{I}(t), \mathbf{S}(t)\right)$ satisfies the following set of differential equations:*

$$\dot{I}_i(t) = -\sum_{j\neq i}^J \lambda_{ij}^{I:k}(\mathbf{I}, \mathbf{S}) \cdot I_i + \sum_{j=1}^J \beta_{ji} I_j S_i + \sum_{j\neq i}^J \lambda_{ji}^{I:k}(\mathbf{I}, \mathbf{S}) \cdot I_j \quad (i = 1, 2, \ldots, J),$$

$$\dot{S}_i(t) = -\sum_{j\neq i}^J \lambda_{ij}^{S:k}(\mathbf{I}, \mathbf{S}) \cdot S_i - \sum_{j=1}^J \beta_{ji} I_j S_i + \sum_{j\neq i}^J \lambda_{ji}^{S:k}(\mathbf{I}, \mathbf{S}) \cdot S_j \quad (i = 1, 2, \ldots, J),$$

where $\lambda_{ij}^{I:k}(\cdot)$ and $\lambda_{ij}^{S:k}(\cdot)$ are valid over the time interval $t \in [T_{k-1}, T_k)$, $k \in \mathbb{Z}_{>0}$. Then

$$\lim_{N\to\infty} \sup_{s\leq t} \left| \tfrac{1}{N}\left(\mathbf{n}^I(s), \mathbf{n}^S(s)\right) - \left(\mathbf{I}(s), \mathbf{S}(s)\right)\right| = 0 \ \textit{ a.s. for all } t > 0.$$

In the equations above, the first terms on the right-hand side represent the movements of vehicles from cluster i to adjacent clusters, while the third terms represent the movement from adjacent clusters to cluster i. The second terms indicate successful transmissions of a V2X message to uninformed vehicles in cluster i from informed vehicles in clusters that are within the communication range.

We provide the proof in the Appendix. When sharp transitions occur between modes, mobility functions of the equations are instantaneously changed, but $\mathbf{x}(t)$ is not reset, so executions of the system are always continuous.

Computation time — As Theorem 1 states, as the number of vehicles approach infinity, actual fraction of informed and uninformed vehicles in all the clusters converge to the corresponding elements of the solution of the differential equations. The computation time taken to solve the set of differential equations does not increase with increase in the number of vehicles. Thus our analytical characterization is computationally tractable regardless of the total number of vehicles in the system. The number of both differential equations and variables are twice the total number of J clusters. An interesting question is how the computation time depends on the number of modes. We answer this considering the case of K traffic lights in a system. The differential equation for a cluster just utilizes the mobility rates in the cluster and in the clusters that feed into the cluster. These mobility rates can be obtained when the states of the corresponding traffic signals are known. If, for example, the signals change states periodically (regardless of any synchronization between them), the states of the corresponding signals at any given time, and hence during a mode, can be obtained in $O(K)$ time. Thus the mobility rates involved in the differential equation for a cluster can be obtained in $O(K)$ time. The total complexity becomes $O(JK)$ because the total number of differential equations are linear to the total number of J clusters.

B. Heterogeneity in V2X communications

We consider a V2X network with two classes of vehicles: *Class 1* consists of certified vehicles who can both transmit and receive, and *Class 2* consists of non-certified vehicles

who can only receive. We also consider a V2X network with infrastructures in the form of RSUs at various locations. Let $n^I_{i:1}$ and $n^I_{i:2}$ respectively be the number of informed vehicles of Classes 1 and 2 in cluster i, and let $n^S_{i:1}$ and $n^S_{i:2}$ respectively be the number of non-informed vehicles of Classes 1 and 2 in cluster i. The $4J$-dimensional vector

$$\begin{aligned}&\big(\mathbf{n}^I_1(t), \mathbf{n}^I_2(t), \mathbf{n}^S_1(t), \mathbf{n}^S_2(t)\big)\\ &= \big(n^I_{1:1}(t), n^I_{2:1}(t), ..., n^I_{J:1}(t); n^I_{1:2}(t), n^I_{2:2}(t), ..., n^I_{J:2}(t);\\ &\quad n^S_{1:1}(t), n^S_{2:1}(t), ..., n^S_{J:1}(t); n^S_{1:2}(t), n^S_{2:2}(t), ..., n^S_{J:2}(t)\big)\end{aligned}$$

represents the state of the system at time t.

Each of the following events corresponds to a state transition. For classes $m = 1, 2$, (1) an informed vehicle of Class m migrates from one cluster to another; (2) a non-informed vehicle of Class m in one cluster receives information through V2V. The events of types (3) and (4) can be obtained from (1) and (2) respectively by replacing 'informed' with 'non-informed', and 'V2V' with 'I2V'.

We now describe the stochastic basis for the evolution of state process $X(t) = \big(\mathbf{n}^I_1(t), \mathbf{n}^I_2(t), \mathbf{n}^S_1(t), \mathbf{n}^S_2(t)\big)$ resulting from these state transitions. We assume that the durations between the following successive successful communications are exponentially distributed: 1) V2V transmission from a vehicle to another it can directly transmit to 2) I2V transmission from an RSU to any vehicle it can directly transmit to. We assume that the sojourn times of the vehicles in each cluster are exponentially distributed. Under these assumptions, the process $X(t)$ constitutes a CTMC. Note that here we consider that I2V communication is unicast, but we will outline how broadcast can be approximated in Section IV. We next provide the rates of these exponential processes.

V2V communication rates — Let the mean of the exponential duration between successive successful transmissions from an informed vehicle of Class 1 in cluster j to a non-informed vehicle in cluster k be $N/\beta_{jk}(\cdot)$. This allows for an exponential backoff duration between communications between the same pair, with its mean depending upon N, the overall number of vehicles. Then $\beta_{jk}(\cdot)/N$ constitutes the communication rate between an informed vehicle of Class 1 in cluster j and a non-informed vehicle in cluster k. $\beta_{jk}(\cdot)$ will in general depend on: (1) nature of the applications; (2) wireless protocol; (3) physical layer communication rates (e.g., fading, interference due to the high density of transmitting vehicles, distance between clusters j, k, etc.). For example, $\beta_{jk}(\cdot) > 0$ only if vehicles in cluster j can directly transmit to those in k, which happens if the clusters are close and there are no blind spots arising from fading; $\beta_{jk}(\cdot) = 0$ otherwise[2] (see an example in Figure3). We expect the inter-cluster communication rate to be lower than the intra-cluster communication rate, i.e., $\beta_{jk}(\cdot) \leq \beta_{jj}(\cdot)$, for all k.

I2V communication rates — Let there be M RSUs, and $\mathcal{S}_m$ be the set of clusters that are within the transmission range of RSU m, for $m \in \{1, \dots M\}$. Then RSU m can directly transmit information to vehicles located in the clusters in set $\mathcal{S}_m$. Let the mean of the exponential duration between successive successful transmissions from RSU m to a non-informed vehicle in any cluster in $\mathcal{S}_m$ be $1/\mu_m$; then μ_m is the rate of such communications.

Mobility rates — Let the mean of the exponential sojourn time of an informed vehicle (a non-informed vehicle, respectively) in cluster j be $1/\lambda^I_j(\cdot)$ ($1/\lambda^S_j(\cdot)$, respectively), and that the vehicle moves to cluster k with a routing probability $p^I_{jk}(\cdot)$ ($p^S_{jk}(\cdot)$, respectively). Then the rate at which an informed vehicle (a non-informed vehicle, respectively) moves from cluster j to k is $\lambda^I_{jk}(\cdot) = p^I_{jk}(\cdot)\lambda^I_j(\cdot)$ ($\lambda^S_{jk}(\cdot) = p^S_{jk}(\cdot)\lambda^S_j(\cdot)$, respectively). Vehicles can directly move from cluster j to k if the clusters are adjacent and traffic rules permit; $p^I_{jk}(\cdot) = 0$ and $p^S_{jk}(\cdot) = 0$, otherwise (see an example in Figure3). The rates $\lambda^I_{jk}(\cdot)$ and $\lambda^S_{jk}(\cdot)$ can depend on the state at that instant, which captures the dependence of vehicular speeds and routing choices on the level of traffic congestion. The rates can also capture that informed and non-informed vehicles may have different mobility patterns due to the influence of the received actionable information.

Mobility and Communication Networks — We had implicitly described the mobility and communication networks through routing probabilities and communication rates between pairs of clusters. The mobility and communication networks overlap but are not identical, equivalently its possible that 1) $p^I_{j,k} = p^S_{j,k} = 0$ but $\beta_{j,k} > 0$, or 2) $p^I_{j,k} > 0, p^S_{j,k} > 0$ but $\beta_{j,k} = 0$. That is, it may not be possible to move directly from one cluster to another, but not communicate between these, the reverse may also happen. For instance, as stated before, physical travel between two adjoining clusters may be infeasible because traffic regulations prohibit vehicles from crossing the median; but vehicles in these may be able to directly communicate between each other as the clusters are in wireless communication range of each other and there are no obstacles between them (see cluster j and k in Figure 3). The versatility of our model lies in its ability to adapt to diverse scenarios. Specifically, our model adeptly handles 1) arbitrary mobility networks characterized by distinct road topologies, and 2) arbitrary communication networks that depend on the proximity of clusters for communication feasibility and local communication conditions.

How to divide the road network into clusters in practice — We divide the road network into segments of a certain size. We typically set this size to the wireless communication range (refer to footnote 2), which corresponds to the natural assumption that the vehicles located within the same cluster can communicate with each other. The communication range depends on the existing communication technology. Next, there are often multiple lanes in different roads in a city and in highways. The lanes in a road segment in which vehicles travel in the same direction are grouped together into one cluster, but if there are lanes in the segment in which vehicles travel in opposite directions, we group them in different clusters. Thus, typically a cluster is chosen as all the lanes in a road segment in which vehicles travel in the same

[2]Wireless communication range is determined by radio capabilities of mobile transmitters and receivers which are usually standardized. If clusters j, k are outside the wireless communication range, $\beta_{j,k} = 0$. But even if clusters j, k are within the wireless communication range of each other, vehicles in cluster j may not be able to communicate with those in cluster k if there are physical obstacles such as tall trees or tall buildings between them and then $\beta_{j,k} = 0$.

direction. For instance, consider the road segment depicted in Figure 3, which comprises of two clusters representing roads in opposite directions. In such a scenario, each cluster may contain multiple lanes that accommodate vehicles traveling in the same direction.

Let $\mathbf{k}$ be the current state, and let $\mathbf{1}_i$ be $4J$-dimensional vector with 1 at its i-th element and 0 elsewhere. The transition rates at which the process jumps from state $\mathbf{k}$ to $\mathbf{k}+\mathbf{h}$ are summarized as follows:

$$q\left(\mathbf{k},\mathbf{k}+\mathbf{h}\right)=\begin{cases}\lambda^I_{jk}\left(\frac{\mathbf{k}}{N}\right)\cdot n^I_{j:1} & \text{if } \mathbf{h}=-\mathbf{1}_j+\mathbf{1}_k,\ j\neq k,\\ \lambda^I_{jk}\left(\frac{\mathbf{k}}{N}\right)\cdot n^I_{j:2} & \text{if } \mathbf{h}=-\mathbf{1}_{J+j}+\mathbf{1}_{J+k},\ j\neq k,\\ \lambda^S_{jk}\left(\frac{\mathbf{k}}{N}\right)\cdot n^S_{j:1} & \text{if } \mathbf{h}=-\mathbf{1}_{2J+j}+\mathbf{1}_{2J+k},\ j\neq k,\\ \lambda^S_{jk}\left(\frac{\mathbf{k}}{N}\right)\cdot n^S_{j:2} & \text{if } \mathbf{h}=-\mathbf{1}_{3J+j}+\mathbf{1}_{3J+k},\ j\neq k,\\ \frac{\beta_{jk}\left(\frac{\mathbf{k}}{N}\right)}{N}n^I_{j:1}n^S_{k:1}+\sum\limits_{m:k\in\mathcal{S}_m}\mu_m n^S_{k:1} & \text{if } \mathbf{h}=\mathbf{1}_k-\mathbf{1}_{2J+k},\\ \frac{\beta_{jk}\left(\frac{\mathbf{k}}{N}\right)}{N}n^I_{j:1}n^S_{k:2}+\sum\limits_{m:k\in\mathcal{S}_m}\mu_m n^S_{k:2} & \text{if } \mathbf{h}=\mathbf{1}_{J+k}-\mathbf{1}_{3J+k},\\ 0 & \text{otherwise.}\end{cases}$$

We now show that the fraction of informed and non-informed vehicles of Classes 1 and 2 in different clusters at any given time t, $X(t)/N$, converges to the output $\mathbf{x}(t)=\big(\mathbf{I}_1(t),\mathbf{I}_2(t),\mathbf{S}_1(t),\mathbf{S}_2(t)\big)$ of the following set of differential equations, as N approaches infinity:

$$\begin{aligned}
\dot{I}_{k:1}(t)=&-\sum_{j\neq k}^{J}\lambda^I_{kj}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{k:1}+\sum_{j=1}^{J}\beta_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:1}S_{k:1}\\
&+\sum_{j\neq k}^{J}\lambda^I_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:1}+\sum_{m:k\in\mathcal{S}_m}\mu_m S_{k:1}\qquad(k=1,2,\ldots,J),\\
\dot{I}_{k:2}(t)=&-\sum_{j\neq k}^{J}\lambda^I_{kj}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{k:2}+\sum_{j=1}^{J}\beta_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:1}S_{k:2}\\
&+\sum_{j\neq k}^{J}\lambda^I_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:2}+\sum_{m:k\in\mathcal{S}_m}\mu_m S_{k:2}\qquad(k=1,2,\ldots,J),\\
\dot{S}_{k:1}(t)=&-\sum_{j\neq k}^{J}\lambda^S_{kj}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)S_{k:1}-\sum_{j=1}^{J}\beta_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:1}S_{k:1}\\
&+\sum_{j\neq k}^{J}\lambda^S_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)S_{j:1}-\sum_{m:k\in\mathcal{S}_m}\mu_m S_{k:1}\qquad(k=1,2,\ldots,J),\\
\dot{S}_{k:2}(t)=&-\sum_{j\neq k}^{J}\lambda^S_{kj}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)S_{k:2}-\sum_{j=1}^{J}\beta_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)I_{j:1}S_{k:2}\\
&+\sum_{j\neq k}^{J}\lambda^S_{jk}(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)S_{j:2}-\sum_{m:k\in\mathcal{S}_m}\mu_m S_{k:2}\qquad(k=1,2,\ldots,J).
\end{aligned}\tag{1}$$

The second and the last terms on the right-hand side of (1) respectively correspond to V2V and I2V communication. The last terms exist only if $M>0$, otherwise the summations are over empty sets. The first and third terms correspond to vehicular movements across clusters. Thus, since V2V communications involve a pair of vehicles, the corresponding terms are quadratic; both I2V and mobility involve only one vehicle, and thus the corresponding terms are linear.

The following theorem guarantees the convergence, and has been proven in the Appendix:

Theorem 2. *Suppose for $j,k=1,2,...,J$, mobility parameters $\lambda^I_{jk}(\cdot):E\to\mathbb{R}$ and $\lambda^S_{jk}(\cdot):E\to\mathbb{R}$, and communication parameters $\beta_{jk}(\cdot):E\to\mathbb{R}$ are bounded and Lipschitz continuous on E. Suppose $\mathbf{x}(0)=\lim\limits_{N\to\infty}X(0)/N$, and also that $\mathbf{x}(t)=\big(\mathbf{I}_1(t),\mathbf{I}_2(t),\mathbf{S}_1(t),\mathbf{S}_2(t)\big)$ satisfies the above differential equations* (1)*. Then*

$$\lim_{N\to\infty}\sup_{s\leq t}\left|\frac{X(t)}{N}-\mathbf{x}(t)\right|=0\qquad\textit{a.s. for all } t>0.$$

In the above theorem, the set E is defined by $E:=\big\{(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2)\ \big|\ I_{i:1}\geq 0, I_{i:2}\geq 0, S_{i:1}\geq 0, S_{i:2}\geq 0, i=1,2,\ldots,J;\ \sum_{i=1}^{J}(I_{i:1}+I_{i:2}+S_{i:1}+S_{i:2})=1\big\}$ with $\big(\mathbf{I}_1,\mathbf{I}_2,\mathbf{S}_1,\mathbf{S}_2\big)=\big(I_{1:1},I_{2:1},\ldots,I_{J:1};\ I_{1:2},I_{2:2},\ldots,I_{J:2};\ S_{1:1},S_{2:1},\ldots,S_{J:1};\ S_{1:2},S_{2:2},\ldots,S_{J:2}\big)$.

By virtue of the above Theorem, for $j=1,2$, in the asymptotic limit of infinite number of vehicles ($N\to\infty$), at any given time t, $\mathbf{I}_j(t)$ ($\mathbf{S}_j(t)$, respectively) represent fraction at t of overall vehicles who are informed (non-informed, respectively) and belong to Class j (lth component of these vectors represent the fraction in cluster l) across clusters.

For easier understanding, we have gone into detail about the sharp time-triggered transitions of mobility (Section II-A) and heterogeneous V2X communication (Section II-B) separately. One may easily generalize the model by incorporating both pulsed traffic mobility and heterogeneous communication in a single model.

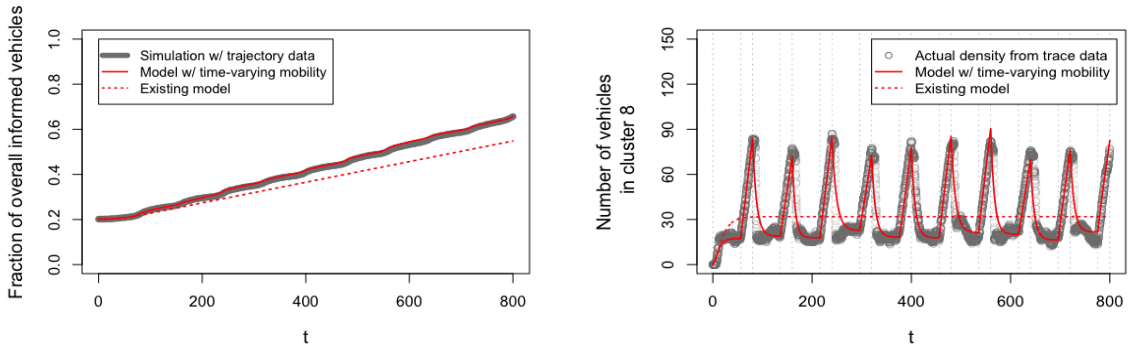
Computation time — The computation time to solve (1) does not depend on the number of vehicles. The number of both differential equations and variables are $4J$ (recall that J is the total number of clusters) and do not increase with the number of RSUs. Thus, the total complexity becomes $O(J)$.

III. Results

In this section, we show that the solution of our model closely matches the simulation result for V2X message flow even when the traffic flow exhibits sharp time-triggered transitions (Sections III-A1 and III-A2). We also show that the match is significantly better than what we observe for the model introduced in prior work [1] which does not incorporate traffic lights. We then use numerical computations from this mathematical framework to answer several questions pertaining that influence the practice of V2X network design and security (Sections III-A2 and III-B).

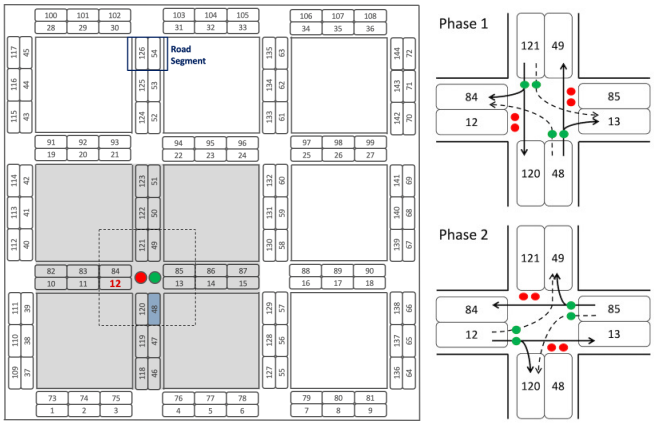
A. Heterogeneity in mobility process - Sharp time-triggered transitions

1) Empirical validation with traffic trace data: We empirically validate the model considering cases in which our modeling assumptions do not hold. Specifically, the theorem in Section II ensures that actual fraction of informed and non-informed vehicles in all the clusters converge to the corresponding elements of the solution of the differential equations when the number of vehicles approaches infinity and vehicles satisfy the exponential sojourn time assumption within each mode. We now consider synthetic trajectory data involving mobility of a finite number of vehicles. The synthetic trajectory data does not satisfy the exponential sojourn time assumption within each mode. This is partly because the traffic lights hold up vehicles for a deterministic duration. Through the empirical validation, we confirm that our mathematical model approximates well the simulation results of V2V message flow involving such trajectory data.



(a) (b)

Fig. 4. (a) Fraction of overall informed vehicles over time. The average deviation over time between the gray points and the dotted red line is 0.3, while the average deviation between the gray points and the solid red line is 0.1. (b) Fraction of vehicles in cluster 8. The gray points represent the actual traffic density based on the synthetic trace data. The average deviation over time between the gray points and the dotted red line is 16.4, while the average deviation between the gray points and the solid red line is 4.9. The solid red line is the solution of the model with time-varying mobility, and the red dotted line is the solution of the previous model [1] assuming a consistent exponential mobility process.



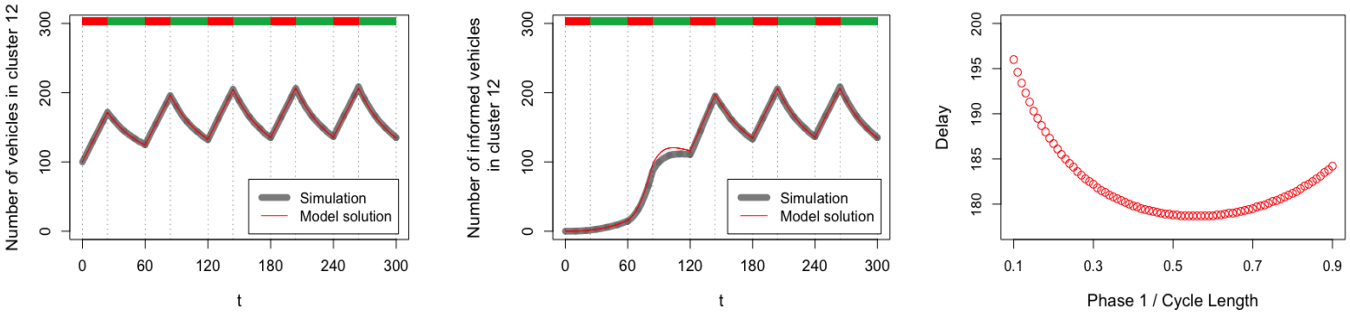
(a) (b)

Fig. 5. (a) Clustered road topology with a traffic signal at one intersection. Here, there are two clusters on the same road segment, one correspond to lanes in which vehicles traverse in a certain direction, another to lanes in which vehicles traverse in the reverse direction. We typically consider the size of a road segment as that equaling wireless communication range. Thus vehicles in the two clusters on the same road segment can communicate with each other. (b) This is an enlarged view of the dotted rectangle around the traffic signal in Figure (a), and represents traffic flow in 2 phases. Broadly, phase 1 corresponds to vehicles moving straight in the vertical direction (e.g., cluster 48 to 49), and phase 2 corresponds to vehicles moving straight in the horizontal direction(e.g., cluster 85 to 84). In addition, vehicles can make left and right turns as indicated by the arrows; for example, vehicles in cluster 48 can move to clusters 13 and 84 during phase 1, and vehicles in cluster 85 can move to clusters 49 and 120 during phase 2.

We use the first 831.7 seconds of actual trace data [24] collected from the U.S. Highway 101 in Los Angeles, California. To reflect the pulse traffic caused by a traffic light, we slightly modify the trace data assuming that there is a traffic light on the road. As in [1], we first divide the road into clusters (Figure 1a), and classify the clusters into three categories; entry clusters A = {10, 12}, exit clusters B = {1, 11}, and study area S = {2, 3, ..., 8}. We assume that approximately 20% of all vehicles (i.e., 402 out of a total of $N = 1993$ vehicles) in the entry clusters already received information before entering study area. Suppose there is a traffic light between clusters 7 and 8. The traffic signal has cycle lengths of 80s, and 30% of the time is spent in red. The signal begins its cycles in green at the start of the simulation. We modify the trajectory of vehicles arriving in cluster 8 during the red light, assuming that these vehicles must be stationary waiting for the signal to turn green. When the light turns green, after the delay during the red light, the vehicle starts to move again according to the trajectory of the original data. Thus, synthetic data with synchronized traffic flow can be generated.

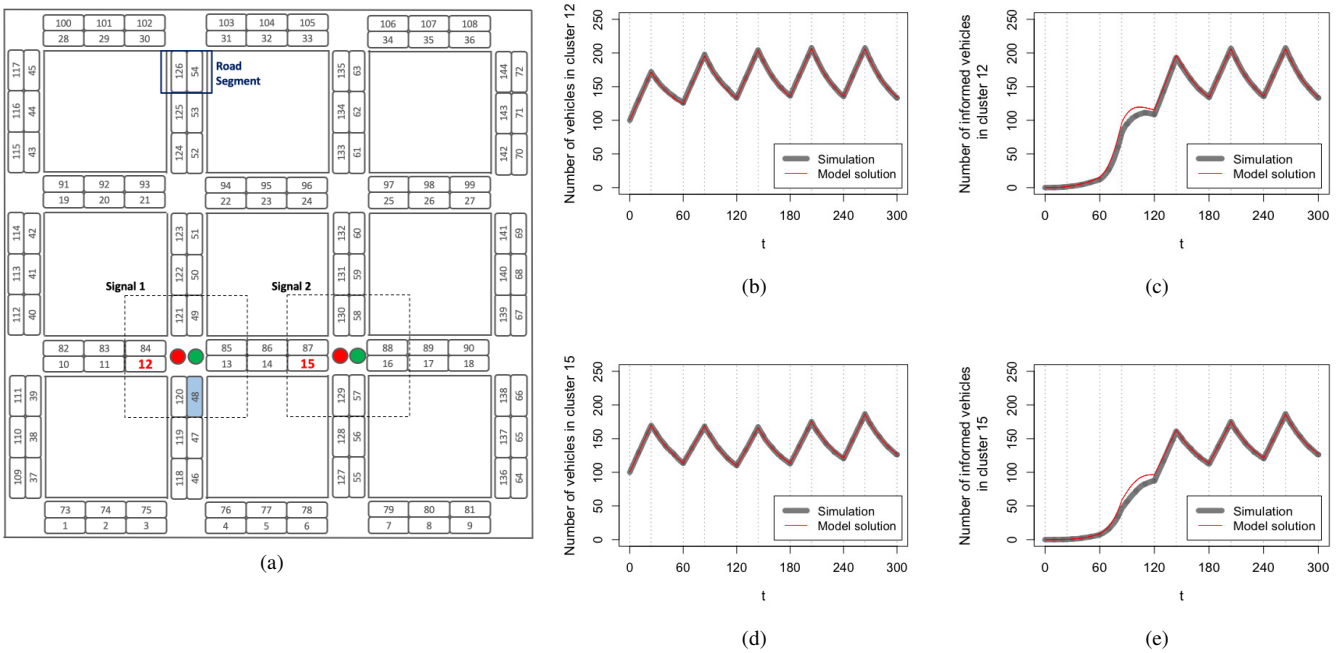
We superimpose the statistical communication process on the trace data (averaged over 100 runs), and then obtain the solutions of both the model with time-varying mobility (introduced in this paper) and the model introduced in prior work [1] (which does not consider sharp time-triggered transitions of mobility patterns). Figures 4a and 4b show that the model with time-varying mobility provides a good match respectively of the fraction of informed vehicles and the overall number of vehicles as functions of time in an example cluster (cluster 8). The match happens despite the fact that the traffic trace corresponds to a finite number of vehicles and does not satisfy the exponential sojourn time assumption within each mode. Figures 4a and 4b also show that the match is significantly better than what we observe for the model introduced in prior work [1].

2) Simulation validation: Using statistical simulations, we investigate the impact of the attributes that do not arise in the synthetic dataset used in Section III-A1: (1) two-dimensional road topology, (2) different durations for each phase of a signal cycle, and (3) multiple traffic lights. We show that the solution of our model closely matches the simulation result for V2V message flow in the presence of one or more traffic lights in two-dimensional grid topology. We also show that the delay in the propagation of V2V messages across the network significantly depends on the relative durations of red and green lights.



(a) (b) (c)

Fig. 6. (a) The number of total vehicles in cluster 12 over time. (b) The number of informed vehicles in cluster 12 over time. The red and green bars on the top indicate the color of the traffic light over time, from the perspective of vehicles located in cluster 12. For (a) and (b), the gray lines represent the average of 100 simulation runs, and the red lines are the model solutions. The traffic signals have cycle lengths of 60 s, and 40% of the cycle time is spent in phase 1. (c) The time delay until 99% of vehicles in the shaded region receive information, which is obtained from the differential equations. This represents how quickly the V2V message spreads to many vehicles.



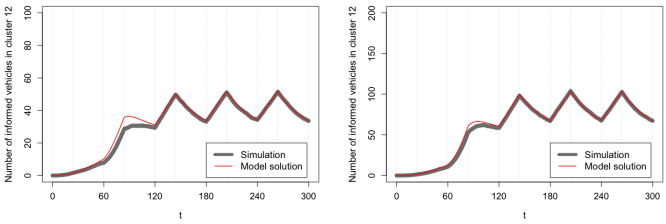
(a) (b) (c) (d) (e)

Fig. 7. (a) Clustered road topology with traffic signals at two intersections. (b) The number of vehicles in cluster 12 over time. (c) The number of informed vehicles in cluster 12 over time. (d) The number of vehicles in cluster 15 over time. (e) The number of informed vehicles in cluster 15 over time. The number of informed vehicles in each clusters is well approximated by model solution. For (b), (c), (d), and (e), the gray line represents the average of 100 simulation runs, and the red line is the model solution. The traffic signals have cycle lengths of 60 s, and 40% of the cycle time is spent in phase 1. Also, signals 1 and 2 are synchronized.

We first consider the presence of traffic signal at one intersection in a two-dimensional grid road topology (Figure 5a). The traffic signal is located at the intersection indicated by the dotted rectangle. The traffic phase design consists of phases 1 and 2 (Figure 5b) and the cycle length is set to 60 seconds. Refer to Figure 5b and its caption to see what the phases mean.

The mobility rates $\lambda_{ij}^{I:k}(\cdot)$ and $\lambda_{ij}^{S:k}(\cdot)$ for clusters $i \in \{12, 48, 85, 121\}$, located right in front of traffic light, are different depending on the phase; the parameters switch along with the lights: they are 0 if the lights are red in mode k, and are at the normal values if the lights are green in mode k. On the other hand, the mobility rates for the clusters $i \in S \backslash \{12, 48, 85, 121\}$ (i.e., those which do not have traffic lights) are constant regardless of the modes.

Let the neighborhood of cluster i be the set of clusters $N_G(i)$ adjacent to cluster i; vehicles in cluster i can only directly move to $j \in N_G(i)$. Let p_{ij} be the probability that vehicles in cluster i move to cluster $j \in N_G(i)$; thus $\sum_{j \in N_G(i)} p_{ij} = 1$. For the clusters $i \notin \{12, 48, 85, 121\}$ that



(a) 25 vehicles per cluster (b) 50 vehicles per cluster

Fig. 8. The number of informed vehicles in cluster 12 over time for (a) 25 vehicles per cluster and (b) 50 vehicles per cluster when there is a traffic signal at one intersection (as illustrated in Figure 5) . The gray lines represent the average of 100 simulation runs and the red lines are the model solutions.

are not directly controlled by the traffic light, we assume the routing probability is uniform (i.e., $p_{ij} = 1/|N_G(i)|$). For the clusters $i \in \{12, 48, 85, 121\}$ located right in front of the traffic lights, p_{ij} is determined by whether it is a left turn or not. In order to reflect the opportunistic left turn in the absence of traffic flow conflicts, we assume that the routing probability for the left turn at a green traffic light is half of the probability of the right turn and straight ahead. The mobility rate set in mode k is

$$\lambda_{ij}^{I:k}(\cdot) = \lambda_{ij}^{S:k}(\cdot) = \begin{cases} p_{ij}\lambda, & i \notin \{12, 48, 85, 121\} \\ p_{ij}\lambda, & \text{if } i \in \{12, 48, 85, 121\}, \text{ and green light,} \\ 0, & \text{if } i \in \{12, 48, 85, 121\}, \text{ and red light} \end{cases}$$

where λ is constant. We set $\lambda = 0.03$. As mentioned in Section II-B and the caption of Figure 5, the size of a road segment equals the wireless communication range and therefore the vehicles in the clusters on the same road segment can communicate with each other. There are two clusters on each road segment. We set $\beta_{ij} = 5$ if clusters i and j are in the same road segment (including $i = j$) and 0 otherwise.

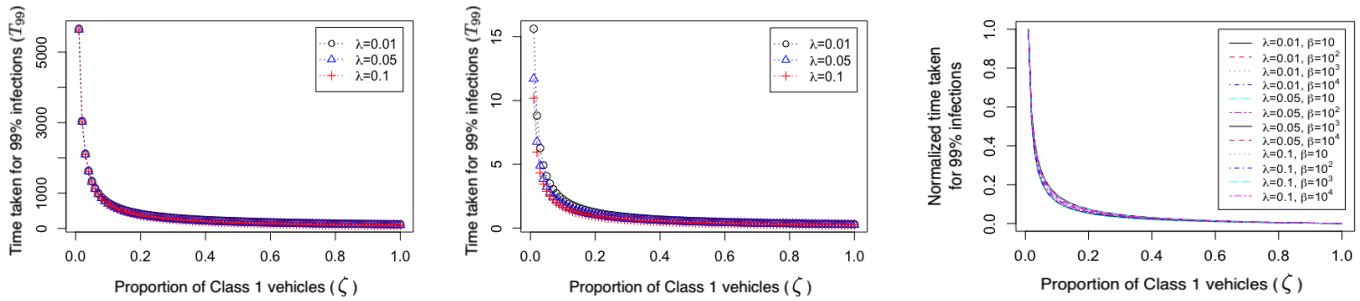
Initially, 14400 vehicles are uniformly distributed in the system (i.e., 100 vehicles per cluster). The shaded area with a traffic light in Figure 5a represents a specific region of interest to investigate information propagation near the traffic signals. We assume that the V2V message initially propagates from 10 vehicles in cluster 48, represented by blue in Figure 5a.

Figure 6a shows that the model solution captures well the traffic flow in cluster 12 despite the burstiness caused by the traffic signal; a signal phase 1 results in traffic accumulating in cluster 12 because vehicles stop at the intersection, followed by the lower traffic volume during a signal phase 2. More importantly, Figure 6b shows that the model captures well the impact of the traffic signal on the spatio-temporal dynamics of message propagation (specifically, the number of informed vehicles in cluster 12 over time). The number of informed vehicles appears smoother in the early stage as there are initially no informed vehicles in the cluster. The smoothness is observed until an informed vehicle from neighboring clusters moves within their communicable distance, subsequently informing the vehicles in cluster 12.

We now show that the delay of the V2V message propagation is significantly affected by the relative durations of the phases in the traffic light, more specifically phases 1 and 2 in Figure 5b. As Figure 6c shows, the time it takes for most vehicles (99%) in the shaded region in Figure 5a to receive information is highly dependent on the time proportion of phase 1.

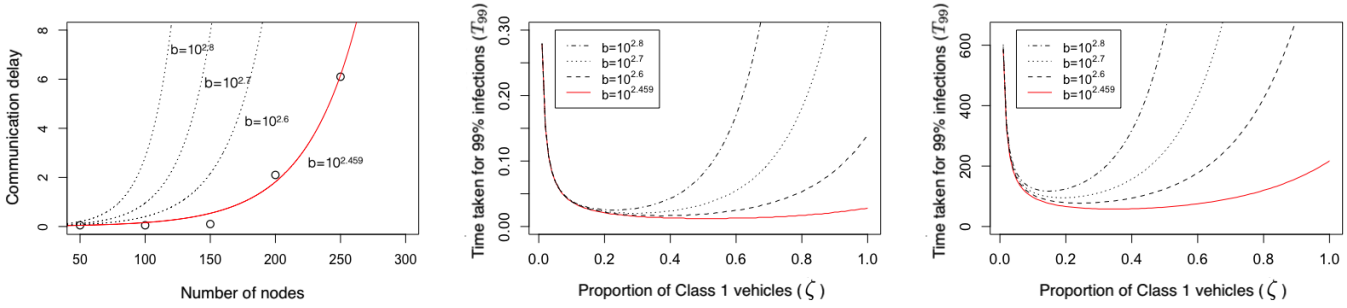
Recall that the mathematical framework can incorporate an arbitrary number of traffic lights and arbitrary coordination patterns between them. We next show how well the model captures the V2V propagation process in the presence of multiple traffic signals. We consider same road topology with traffic signals at two intersections (Figure 7a). The simulation setup is the same as the previous one. The only difference is that with the addition of another traffic light (Signal 2 in Figure 7a), the mobility rate near the added traffic light also changes depending on the phase. We assume that the design of the traffic phase for the added traffic signal 2 is the same as the design of the phase for the original traffic signal 1, so both traffic lights have identical traffic phase design, as illustrated in Figure 5b. We also assume that both signals are synchronized, so the mode changes at the same time when the phases of both traffic lights are changed. Figures 7b and 7d show that the model solution captures the pulsed traffic movement caused by the traffic signals well. Further, Figures 7c and 7e show that the temporal change in the number of informed vehicles in the particular locations (clusters 12 and 15, respectively) is also closely approximated by the model solution. As in Figure 6b, the number of informed vehicles appears smoother in the early stage as there are initially no informed vehicles in the cluster.

We now demonstrate that our model satisfactorily captures the influence of the traffic signal on the spatio-temporal dynamics of V2X message propagation even when the number of vehicles per cluster is relatively small (Figure 8). We consider the same conditions as depicted in Figure 5, where the traffic flow exhibits sharp time-triggered transitions. Even with a small number of vehicles per cluster, specifically 25, 50, solution of the differential equations (model solution) closely approximates the simulation results for the number of informed vehicles in different clusters. We plot the simulation result (the gray curve) and model solution (the red curve) for the number of informed vehicles in cluster 12 over time. In both cases, the two curves exhibit their maximum discrepancies during the second cycle of the traffic signal, i.e., $t \in [61, 120]$, with average percentage discrepancy as (a) 16.9% and (b) 11.2%, respectively. However, the average percentage discrepancy between the third and fifth cycles, i.e., $t \in [121, 300]$, drop to (a) 1.2% and (b) 0.9%, respectively. Thus overall the percentage discrepancies are acceptable. Note that the average percentage discrepancy over time between t_1 and t_2 is defined as $\frac{1}{t_2-t_1+1}\sum_{t=t_1}^{t_2}\frac{|S_t-M_t|}{M_t}$, where S_t and M_t are simulation result and model solution at time t, respectively. Note also that the percentage discrepancy is not a meaningful metric when the number of informed vehicles is very small, as then the denominator is so small that even minor discrepancies lead to high percentage discrepancies. We therefore do not consider the percentage discrepancies during the first cycle as values of both simulation and model solution are very small in this early phase.



(a) (b) (c)

Fig. 9. T_{99} as a function of $\zeta \in [0.01, 1]$ for various values of mobility parameter $\lambda = 0.01, 0.05$, and 0.1, and communication parameters (a) $\beta = 10$ and (b) $\beta = 10^4$. (c) We normalize T_{99} to lie in the range $[0, 1]$ considering every combination of $\beta \in \{10, 10^2, 10^3, 10^4\}$ and $\lambda \in \{0.01, 0.05, 0.1\}$. The maximum difference between the uppermost and lowermost curves in (c) is 0.0583.



(a) Communication delay vs. density (b) $\beta = 10^{5.92}$ (c) $\beta = 10^2$

Fig. 10. (a) The points represent a rough estimate from figure 9 of Eichler's work [26]. The red solid line represents the fitted line to the points ($\beta = 10^{5.92}$, $b = 10^{2.459}$ in (2)), and the dotted lines represent the delay for larger values of b. (b) T_{99} as a function of $\zeta \in [0.01, 1]$ for $\beta = 10^{5.92}$ and various values of parameter b. (c) T_{99} as a function of $\zeta \in [0.01, 1]$ for the smaller value of $\beta = 10^2$ and the same values of parameter b.

B. Heterogeneity in V2X communications

In this section, we use the differential equation formulation to answer the questions that heterogeneity poses. Specifically, we investigate what percentage of certified vehicles would suffice towards ensuring the desired propagation of the information in the absence of RSUs (1) when the transmitting vehicles use the same communication rates everywhere, and (2) when communication rates depend on local traffic density, and associated interference. We next revisit (1) and (2) in presence of RSUs. These three subsections respectively relate to the three questions posed in the Introduction. The answers we obtain reveal several attributes of V2X systems that would influence practice. We demonstrate that (1) speed of propagation of V2X information is maximized when only a small proportion of vehicles are allowed to transmit; increasing the transmitting fraction beyond this threshold either provides diminishing return or due to interference decreases the speed of information propagation, (2) information spreads much faster through V2V communication than through I2V, considering systems with only one or both of these; specifically, with even a small proportion of vehicles transmitting, the differential impact of increasing the number of RSUs on propagation speed is marginal, despite using a parameter setting that favors I2V over V2V.

We use the performance metric *the time it takes for 99% of all vehicles in the system to receive information*, which we denote by T_{99}. Note that the unit of time used here is seconds. Investigating how T_{99} depends on ζ, *the proportion of Class 1 vehicles in the system*, we answer the three questions listed in Section I sequentially in each subsection of the current section..

We consider a two-dimensional grid road topology with bidirectional traffic (Figure 3). As depicted by the blue circle in Figure 3, at every intersection, vehicles travel straight, turn right, or turn left. We assume here that vehicles move to the neighboring cluster with equal probability at every intersection. The mobility rate set in this model is $\lambda^I_{ij}(\cdot) = \lambda^S_{ij}(\cdot) = p_{ij}\lambda$, where p_{ij} is the probability of vehicles in cluster i to move to cluster j, and λ is a constant. Assuming a standard V2V transmission range of 300 m, we set the length of one cluster to about 300 m. We consider $\lambda \in [0.01, 0.10]$, which corresponds to average vehicular velocities between 10.8 km/h and 108 km/h. Suppose also that a road segment consists of two clusters corresponding to opposite directional roads (Figure 3). We allow inter-cluster communication between

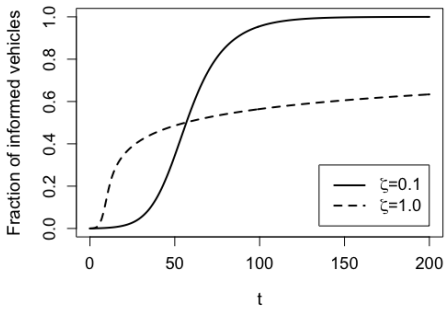
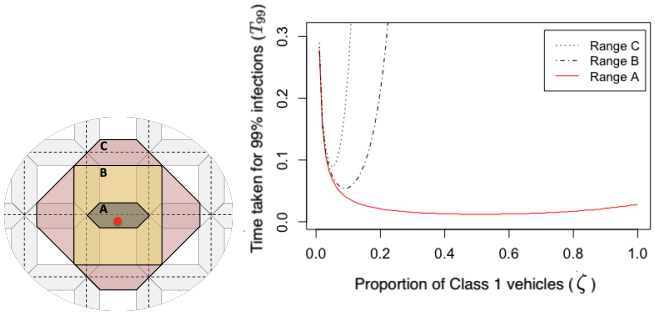


Fig. 11. The solid (dotted, respectively) line represents the fraction of overall informed vehicles over time for $\zeta = 0.1$ ($\zeta = 1.0$, respectively).



(a) (b)

Fig. 12. (a) Three different interference ranges from the perspective of the node indicated by the red circle. (b) T_{99} as a function of $\zeta \in [0.01, 1]$ for the respective ranges with $\beta = 10^{5.92}$ and $b = 10^{2.459}$.

vehicles in clusters within the same road segment, at the same rate as the intra-cluster communication.

In Sections III-B1 and III-B2 we consider that there is no RSU, i.e., $M = 0$. In Section III-B3, there are RSUs, and M is varied between 1 and 8.

Initially, we assume that there are 100 vehicles per cluster. Given that there are 120 clusters in total, the total number of vehicles in the entire topology $N = 12000$. Initially, a single Class 1 vehicle in the red cluster (Figure 3) is informed.

1) Uniform V2V communication, No RSU: We consider V2V communications with uniform rate, i.e., $\beta_{ji}(\cdot) = \beta$, for all j, if $i = j$, or i and j represent clusters within the same road segment; $\beta_{ji}(\cdot) = 0$ for all other i, j pairs (refer to Figure 3 for some illustrations). We consider parameter β in the range $[10, 10^4]$. Since $N = 12000$, the expected duration N/β between successive successful communications between pairs of vehicles in the same cluster or in clusters within the same road segment belongs to a range of 1.2 seconds to 20 minutes; this corresponds to applications that are less time-critical.

Figures 9a and 9b reveal that T_{99} is a monotonically decreasing function of ζ. This happens because increasing ζ increases the proportion of vehicles that are allowed to transmit information, thereby increasing the speed of the information propagation. Secondly, as we increase ζ, initially T_{99} decreases sharply, subsequently, T_{99} decreases very slowly.

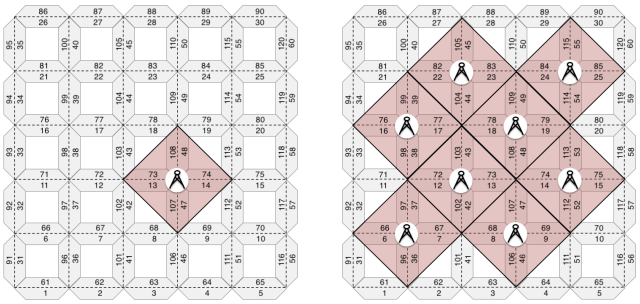
The figures also show that the absolute value of T_{99} substantially varies with β and vehicular speed (λ). But, normalizing T_{99} for $\zeta \in [0.01, 1]$ to lie in the range $[0, 1]$, we notice that the normalized curves of T_{99} for different values of β, λ are near identical (see Figure 9c). In fact, up to a certain *threshold* value of ζ, the normalized T_{99} reduces rapidly to 0.15 and subsequently decreases very slowly with increase in ζ. We notice that this threshold point is almost invariable to changes in β and λ (the threshold points are in the range 0.07 – 0.09). Thus, regardless of the exact choice of β and λ, with increase in ζ, T_{99} decreases to a small percentage of its maximum value, by $\zeta \leq 0.09$; increasing ζ further provides diminishing return in terms of speed of propagation of the message. Thus, it is sufficient to have only 9% of the vehicles transmit, and have the rest only receive.

2) Interference-limited local density-dependent V2V communication, No RSU: We now consider that interference due to high local density of transmitting vehicles throttles V2V communication rates and investigate how the results in Section III-B1 change. We demonstrate that the speed of propagation of information is maximized for a small ζ (i.e., a small percentage of transmitting vehicles) and rapidly decreases with increase in ζ beyond a certain threshold owing to increase in interference level.

The system model is identical to Section III-B1, except that, to capture interference, the uniform communication rate parameter β is substituted by a local vehicular density-dependent function which we describe next. Let $\mathcal{R}_i$ denote the set of clusters whose vehicles' transmissions interfere with the reception by vehicles in cluster i, including cluster i itself. Then $\sum_{k \in \mathcal{R}_i} I_{k:1}$ represents the fraction of overall transmitting vehicles that are currently present in $\mathcal{R}_i$, and therefore the vehicular population share whose transmissions interfere with the reception by vehicles in cluster i. Now, for clusters i, j in the same road segment including when $j = i$,

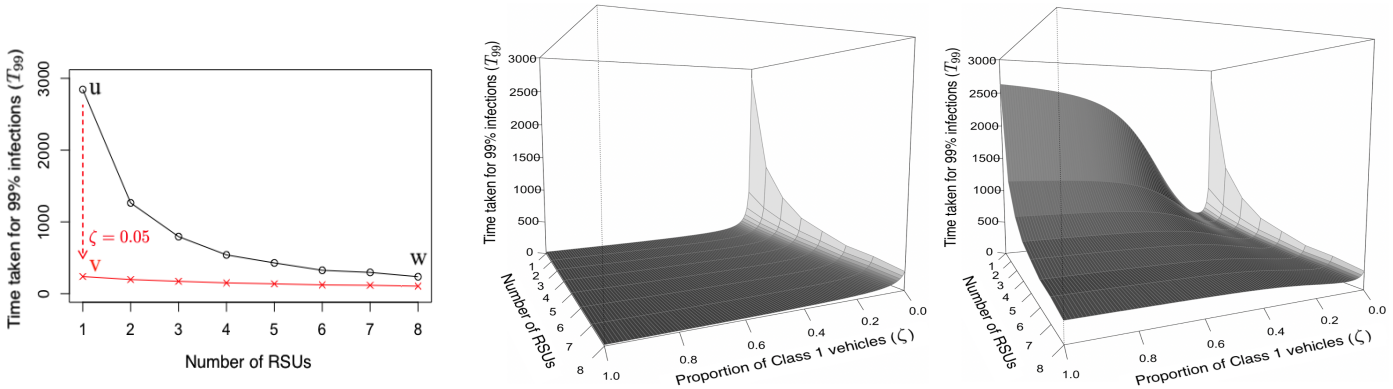
$$\beta_{ji}(\cdot) = \beta \cdot \left(1 - \sum_{k \in \mathcal{R}_i} I_{k:1}\right)^b \tag{2}$$

where $\beta \geq 0$ and $b \geq 1$; β corresponds to the communication



(a) 1 RSU (b) 8 RSUs

Fig. 13. One scenario for distributing RSUs. (a) One RSU and its communication range. (b) Eight RSUs and the area covered by the RSUs.



(a) (b) (c)

Fig. 14. Figure (a) shows T_{99} as a function of the number of RSUs when $\zeta = 0$ (circles) and $\zeta = 0.05$ (crosses) under the uniform communication model. The points u and v respectively represent the cases that there is 1 RSU and 1) no V2V (i.e., $\zeta = 0$) 2) 5% vehicles are in Class 1 (i.e., $\zeta = 0.05$). The point w represents the case that there are 8 RSUs and no V2V (i.e., $\zeta = 0$). The values of vertical axis are 2844.4 for u, 239 for v, and 236 for w. Figures (b) and (c) respectively show the comprehensive impact of different combinations of $\zeta \in [0, 1]$ and the number of RSUs on T_{99} (b) under the uniform V2V communication model, and (c) the density-dependent V2V communication model.

parameter in the absence of interference. For a given β, b determines the sensitivity of communication rate to the interference level associated with the local density of transmitting vehicles in the interference range of each cluster.

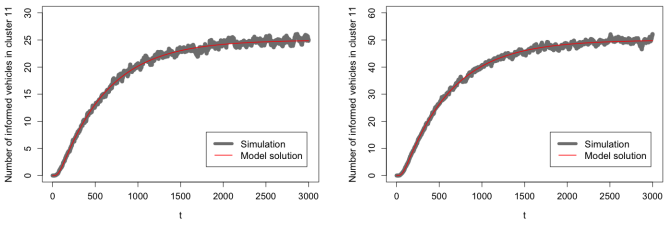
The above functional form of dependence of the communication rate on the local vehicular density has been inspired by the end-to-end delays in the IEEE 802.11p standard obtained through analysis and simulation in [26]. The points in Figure 10a are rough estimates of the relation between the end-to-end delay and the number of nodes in the system based on the simulations presented in [26][3]. The solid line in Figure 10a represents the fitted line to the points. We estimate $\beta = 10^{5.92}$ and $b = 10^{2.459}$, for the value of N we use, by choosing β, b to minimize the least squares deviation from the delay data reported in [26][4]. Given that the parameter estimates are an artifact of the specific setting used in [26] which is not identical to ours, we consider a range of values of b, $\beta, \mathcal{R}_i$. For example, the dotted lines in 10a represent the delay for larger values of b, which increases faster than the line obtained from data points given in [26]. We set the mobility parameter λ to be 0.05.

Impact of ζ — We observe that increasing ζ beyond a certain value slows the spread of information. We first use β, b represented by the lines in Figure 10a, and set $\mathcal{R}_i$ to consist of clusters i and j, where i, j correspond to the same road segment, i.e., interference range is the same as the transmission range. Figures 10b and 10c show that when ζ is increased, there is a sharp drop in T_{99} at the beginning, followed by a reasonably flat phase and an increase beyond a certain value of ζ, $\zeta_{\max}$. This happens because a greater number of transmitting vehicles increases interference, which decreases the communication rate. Also, the larger the b value, the smaller is $\zeta_{\max}$ (0.57 for $b = 10^{2.459}$, 0.38 for $b = 10^{2.6}$, 0.31 for $b = 10^{2.7}$, and 0.25 for $b = 10^{2.8}$ in Figure 10b; and 0.39 for $b = 10^{2.459}$, 0.25 for $b = 10^{2.6}$, 0.2 for $b = 10^{2.7}$, and 0.15 for $b = 10^{2.8}$ in Figure 10c). Thus, $\zeta_{\max}$ decreases with increase in b, the sensitivity of the communication rate function to interference. Figures 10b and 10c differ only in the scale of the vertical axis; this difference arises because, as expected, T_{99} increases with decrease in β, the maximum value of the communication rate.

We now study how the overall fraction of informed vehicles changes with time, considering $\beta = 10^2, b = 10^{2.8}$ as an example. As Figure 11 shows, at an early stage, information propagates faster when $\zeta = 1.0$ than when $\zeta = 0.1$, but the observation is reversed from a certain time onward. This is because the maximum number of transmitting vehicles is limited to 10% of the total number of vehicles when $\zeta = 0.1$, while all vehicles can transmit the information upon receipt of the information when $\zeta = 1.0$. At an early stage, the number of informed vehicles is low, thus interference is low for all values of ζ, and information propagates rapidly in cases in which a larger number of vehicles transmit (i.e., $\zeta = 1.0$). But, as more vehicles become informed, a larger number of transmissions

[3]In [26], application messages are classified into different access classes (ACs), of which AC0 has the lowest priority. [26] showed that the average delay for the low-priority data packets increased exponentially as node density increased. We use the delays for AC0 reported in Figure 5 of [26]. The data point corresponding to 300 nodes on the x-axis has been excluded from the fitting. Excluding this point leads to a fit with a relatively stable phase until 150 nodes, followed by a faster increase. This better represents observations in communication theory, wherein vehicular density only has a marginal impact on the communication delay up to the point when the wireless channel gets saturated, beyond which it severely throttles the delay.

[4]The value of b we obtain, ie $b = 10^{2.459}$, comes across as rather high. This raises the concern that if the communication rate would approach 0 rapidly even when the number of transmitting vehicles in a cluster increases slightly. This concern does not materialize for the following reason. Referring to the functional form in Equation 2), we note that the communication rate is expressed as a function of $I_{k:1}$, which is the number of transmitting (equivalently informed) Class 1 vehicles in cluster k, divided by the total number of vehicles N in the entire system. Since the denominator, which represents the total number of vehicles, is much larger than the number of Class 1 vehicles in a particular cluster k, the value of $I_{k:1}$ is small. For instance, in this paper's scenario with $N = 12000$ total vehicles in the system, adding one Class 1 vehicle to the cluster k results in only a $1/12000$ increase in the value of $I_{k:1}$. Consequently, despite the seemingly large value of b, the communication rate remains at a decent level and does not approach zero rapidly (unless the vehicular density is really high, but in this case the communication rate becomes low even in practice).



(a) 25 vehicles per cluster (b) 50 vehicles per cluster

Fig. 15. The number of informed vehicles in cluster 11 over time for (a) 25 vehicles per cluster and (b) 50 vehicles per cluster when there is 1 RSU and no V2V (as illustrated in Figure 13a). The gray lines represent the average of 100 simulation runs and the red lines are the model solutions.

increase interference and throttles communication rates; thus, propagation speed drastically reduces when $\zeta = 1.0$ as compared to when $\zeta = 0.1$.

Impact of $\mathcal{R}_i$ — We consider three different interference ranges, A, B, C of progressively increasing sizes, leading to progressively increasing $\mathcal{R}_i$s (Figure 12a). The range A is the same as the communication range, and the ranges B and C respectively cover road segments corresponding to five and nine times the communication range. For all the ranges, we use the fitted model of the communication rate (the solid red line in Figure 10a).

Figure 12b shows that the larger the interference range, the smaller is $\zeta_{\max}$ (0.57 for range A, 0.09 for range B, and 0.05 for range C). This is because the larger the interference range, the larger the vehicular population share that interferes with a given transmission, which exacerbates the impact of interference and in effect ensures that the spread of the message starts slowing down at a smaller value of ζ.

To summarize Sections III-B1 and III-B2, increasing ζ beyond certain threshold values yields diminishing gain or worsens the speed of V2V propagation of information.

3) Impact of transmission from RSUs: We investigate the relative efficacy of I2V over V2V, and which enhances the spread - greater share of transmitting vehicles (ζ) or a larger number of RSUs (M). We demonstrate that, regardless of the number of RSUs and regardless of whether we account for interference, as before, increasing ζ beyond a certain point yields only marginal benefit or slows down the spread of information. Also, with even a small ζ, the differential impact of increasing the number of RSUs on propagation speed is marginal. Thus, V2V is more effective than I2V in spreading V2X messages.

We assume that each RSU is located at an intersection, excluding the outer boundary; but not every intersection has an RSU. When there are multiple RSUs, we place them so that the areas covered by each RSU do not overlap (Figure 13); thus, $\mathcal{S}_m$s, $m = 1, \ldots, M$ are mutually exclusive. We set the mobility parameter $\lambda = 0.05$, and the V2V communication parameter $\beta = 40$ for both uniform and density-dependent communication models. We choose the I2V communication rate $\mu_m = 10$ for all the deployed RSUs, $m = 1, \ldots, M$. Thus, for $N = 12,000$, the average I2V and V2V communication delays are respectively 0.1 sec and 5 min (we later explain the significance of this imbalance on our results). We average the results over 10 random geographic locations for each RSU for $1-7$ RSUs. For 8 RSUs, only 2 configurations are possible; we average our results over both. We consider that no vehicle is initially informed, and initially information enters the vehicles from the RSUs.

We first consider a scenario with only I2V (i.e., $\zeta = 0$, thus there is no V2V). In this case, T_{99} noticeably decreases as the number of RSUs increases as expected (see black circles in Figure 14a). Then we additionally consider V2V, with uniform V2V communication model. We observe that with 1 RSU, even $\zeta = 0.05$ substantially decreases T_{99} compared to $\zeta = 0$ (see sharp decrease from point u to v in Figure 14a). This topology can have no more than 8 RSUs. And, T_{99} with 1 RSU, $\zeta = 0.05$, is comparable to T_{99} for 8 RSUs, $\zeta = 0$ (see points v to w in Figure 14a). Thus, despite using a parameter setting that favors I2V over V2V (recall that I2V communications have much lower delay than V2V ones), the differential impact of increasing the number of RSUs is marginal even with a small proportion of vehicles transmitting (see red crosses in Figure 14a). This suggests that V2V communication enhances the spread of information much more than I2V. This happens as a vehicle encounters much greater number of other vehicles in its communication range than an RSU, as RSUs are fixed while vehicles move. The phenomenon can be seen more clearly in Figure 14b that captures the comprehensive impact of different combinations of ζ and the number of RSUs on information propagation.

We now investigate if and how the previously observed phenomena change when we consider the density-dependent V2V communication model. We choose $b = 10^{2.459}$ in (2), and consider the interference range B represented in Figure 12a. When there is 1 RSU, as we have observed in Section III-B2, there is a sharp drop in T_{99} at the beginning as ζ increases, followed by an increase in T_{99} beyond a certain value of ζ, $\zeta_{\max}$ (see Figure 14c). Thus, the threshold phenomenon continues to hold. But, as the number of RSUs increases, the rate of increase of T_{99} beyond $\zeta_{\max}$ slows down because the additional I2V communication compensates for the reduction in the V2V communication rates due to interference (fewer vehicles need to receive the information through V2V). Nonetheless, we have observed that as in Section III-B2, the $\zeta_{\max} \approx 0.07$, regardless of the number of RSUs in this case. Thus, again, the information spread is maximized if only a small percentage of vehicles are allowed to transmit. Figure 14c also shows that the differential impact of increasing the number of RSUs is marginal if $\zeta \approx \zeta_{\max}$.

Additionally, we demonstrate that the model satisfactorily captures the V2X message propagation in presence of RSUs for a limited number of vehicles, namely 25, 50 per cluster (Figure 15). We consider a single RSU and no V2V communication as in Figure 13a. We plot the number of informed vehicles in cluster 11 over time, as obtained from the model (red line) and from simulation (gray line). The average and maximum percentage discrepancies between the two over time are respectively (a) 3.0% and (b) 2.4%.

IV. Conclusion and Discussion

We have provided a mathematical framework for characterizing the dynamics of information propagation in V2X system under pulsed traffic as well as communication heterogeneity comprising of (1) different V2V commutation capabilities, (2) different wireless communication conditions, and (3) both V2V and I2V. The computation time for the framework does not increase in the number of vehicles and RSUs, but only increases linearly in the network size (number of clusters) and the number of traffic lights.

A. Implications of our findings on practice

Numerical computations using this framework reveal several attributes of V2X systems that would influence practice. (1) Considering various communication and mobility conditions, we show that, speed of propagation of V2X information is maximized when only a small proportion of vehicles are allowed to transmit; increasing the transmitting fraction beyond this threshold either provides diminishing return or due to interference slows down information propagation. Thus the resilience of V2X to cyber-attacks may be enhanced, by restricting transmission capabilities to a small number of authenticated vehicles, without reducing efficacy of spread of information. (2) We demonstrate that information spreads much faster through V2V communication than through I2V, considering systems with only one or both of these. Specifically, with even a small proportion of vehicles transmitting, the differential impact of increasing the number of RSUs on propagation speed is marginal, despite using a parameter setting that favors I2V over V2V. This will have a significant bearing on the practices of 1) V2X network design, as deploying and managing RSUs may incur significant expenses, and 2) V2X security as securing I2V from RSUs may be easier than securing V2V communications. The silver lining with respect to 2) is that it suffices to have only a small fraction of vehicles transmit, and therefore securing their transmissions may not be as challenging.

B. Discussion and Future Research

Mobility rate for non-exponential sojourn time — The mathematical guarantees for our model holds as the number of vehicles approach infinity and the sojourn times of vehicles in various clusters, intervals between communication of V2V messages are exponentially distributed[5]. An interesting direction for future research is to obtain similar or even somewhat weaker mathematical guarantees, when the above assumptions do not hold. The sojourn times of the vehicles in individual clusters will not in general be exponentially distributed owing to the vehicular queues that build up when for example the vehicles stop at red lights or due to heavy traffic congestion or slow down due to the latter. One idea for generalization of the mathematical guarantees is to choose vehicular mobility rates that appropriately depend on the congestion levels in the clusters, and mathematically prove that the average behavior is captured.

General heterogeneity of vehicles — We have considered the impact of communication heterogeneity so far by considering two classes of vehicles, one of which both transmit and receive, and the other only receive, and also both V2V and I2V. More fine-grained heterogeneity can be accommodated by generalizing to multiple classes of vehicles corresponding for example to different ranges of communication rates. The classifications may be determined by hard system constraints or by strategic choice of communication rates. The classifications may also correspond to different mobility characteristics, which would in general be different for different types of vehicles (e.g., shared ride, personal vehicles, buses, and freight trucks, etc.), and their different stages of automation. For example, our recent study [1] provided an outline of a mathematical framework for considering vehicles classified by their destinations. By integrating these frameworks, we can accommodate a general classification of vehicles pertaining to different mobility and communication capabilities and choices, and possibly transition between the categories governed by the strategic choices. In this case, $I_{j:k}(t)(S_{j:k}(t))$ would represent fraction of overall vehicles which are informed (non-informed, respectively), and are in cluster j and category k.

Broadcast from RSU — We have so far considered unicast (one-to-one) I2V communications between RSUs and vehicles. In practice, a RSU can broadcast an information to all vehicles in its communication range through periodic one-to-many communications. We can approximately capture this scenario by introducing time-dependent I2V communication rates $\mu_j(t)$s. The overall time interval will be divided into broadcast intervals; throughout a broadcast interval, except at a small sub-interval at the end, a RSU m will not transmit anything and $\mu_m(t) = 0$; in the small sub-interval at the end, the RSU will transmit at a very high rate (i.e, high $\mu_j(t)$), due to which with a very high probability all non-informed vehicles within its communication range will receive the message. Thus, the unicasts in the concluding small sub-interval that inform all vehicles in the communication range mimic an one-to-many (broadcast) communication at the end of the designated broadcast intervals. We can generalize the time-dependent mobility rates that switch discontinuously to different values over time that we introduced to model the impact of traffic signals to incorporate the time-varying I2V communication rates switching discontinuously over time, which mimic periodic broadcast.

Limitation on the number of vehicles per cluster — There exists hard limits on the number of vehicles in a road segment due to physical capacity constraints. Our model does not consider limits on occupancy in road segments. But, still in all our numerical computations, we have seen that the occupancy in the roads are well within bounds imposed by realistic physical capacity constraints even when we do not explicitly impose such hard limits. For example, in the two example

[5]The exponential process assumption have been made before in literature of transportation research, e.g. [1], [27]. The paper [1] is a prior research of the authors of this paper and have been published in a premier journal of the transportation research community. The exponential assumption is equivalent to assuming that the the distribution of vehicles across time and space is a Markov process. This Markovian assumption has been made and verified with empirical data in [27]. Markov models have in fact been extensively utilized in the transportation research community, e.g., to estimate freeway travel time in both routine and perturbed states as in [28]–[31].

clusters in front of the traffic lights considered in Figure 7a, the number of vehicles varies between approximately 100 and 200 (Figure 7b and 7d). Recall that we have assumed the cluster length of 300m which is V2V transmission range. If we approximate the length of a car by 5m, this would allow for 60 vehicles per lane. Several roads have 4 or more lanes. Thus, the maximum value we observe in this case is lower than the capacity (240 vehicles). Note that we have considered an initial density of 100 vehicles per cluster, which does not represent sparse traffic condition. We have chosen this number by considering a common scenario of slightly over 40% traffic capacity on a four-lane road. Since the clusters are in front of the traffic lights, traffic tends to accumulate in these. Thus, even in clusters where traffic tends to accumulate, and when initial density is not low, the maximum numbers of vehicles in the clusters are below the road capacity. Nonetheless, incorporating such constraints is important from an academic point of view. The resulting models will involve a system of constrained differential equations, which constitutes an interesting direction of future research.

REFERENCES

[1] J. Kim, S. Sarkar, S. S. Venkatesh, M. S. Ryerson, and D. Starobinski, "An epidemiological diffusion framework for vehicular messaging in general transportation networks," Transportation Research Part B: Methodological, vol. 131, pp. 160–190, 2020.

[2] M. Ferreira, R. Fernandes, H. Conceição, W. Viriyasitavat, and O. K. Tonguz, "Self-organized traffic control," in Proceedings of the seventh ACM international workshop on VehiculAr InterNETworking, 2010, pp. 85–90.

[3] New York City Department of Transportation (NYCDOT). Infrastructure: Traffic Signals. (Accessed Jan 21, 2020). [Online]. Available: https://www1.nyc.gov/html/dot/html/infrastructure/signals.shtml

[4] J. Li, H. Lu, and M. Guizani, "Acpn: A novel authentication framework with conditional privacy-preservation and non-repudiation for vanets," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 4, pp. 938–948, 2014.

[5] K. P. Laberteaux, J. J. Haas, and Y.-C. Hu, "Security certificate revocation list distribution for VANET," in VANET '08 - Proceedings of the 5th ACM International Workshop on VehiculAr Inter-NETworking. ACM, 2008, pp. 88–89.

[6] J. J. Haas, Y.-C. Hu, and K. P. Laberteaux, "Efficient certificate revocation list organization and distribution," IEEE Journal on Selected Areas in Communications, vol. 29, no. 3, pp. 595–604, 2011.

[7] H. Zhu, R. Lu, X. Shen, and X. Lin, "Security in service-oriented vehicular networks," IEEE Wireless Communications, vol. 16, no. 4, pp. 16–22, 2009.

[8] S. K. Dhurandher, M. S. Obaidat, A. Jaiswal, A. Tiwari, and A. Tyagi, "Vehicular security through reputation and plausibility checks," IEEE Systems Journal, vol. 8, no. 2, pp. 384–394, 2013.

[9] L. C. Bento, R. Parafita, and U. Nunes, "Intelligent traffic management at intersections supported by v2v and v2i communications," in 2012 15th international IEEE conference on intelligent transportation systems. IEEE, 2012, pp. 1495–1502.

[10] G. De Nunzio, C. C. De Wit, P. Moulin, and D. Di Domenico, "Eco-driving in urban traffic networks using traffic signals information," International Journal of Robust and Nonlinear Control, vol. 26, no. 6, pp. 1307–1324, 2016.

[11] M. A. S. Kamal, S. Taguchi, and T. Yoshimura, "Intersection vehicle cooperative eco-driving in the context of partially connected vehicle environment," in 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, 2015, pp. 1261–1266.

[12] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," IEEE Transactions on Vehicular Technology, vol. 68, no. 2, pp. 1243–1253, 2019.

[13] T. Tan, F. Bao, Y. Deng, A. Jin, Q. Dai, and J. Wang, "Cooperative deep reinforcement learning for large-scale traffic grid signal control," IEEE transactions on cybernetics, 2019.

[14] Y. H. Kim, S. Peeta, and X. He, "Modeling the information flow propagation wave under vehicle-to-vehicle communications," Transportation Research Part C: Emerging Technologies, vol. 85, pp. 377–395, 2017.

[15] J. Wang, Y. H. Kim, X. He, and S. Peeta, "Analytical model for information flow propagation wave under an information relay control strategy in a congested vehicle-to-vehicle communication environment," Transportation Research Part C: Emerging Technologies, vol. 94, pp. 1–18, 2018.

[16] H. Wu, R. M. Fujimoto, G. F. Riley, and M. Hunter, "Spatial propagation of information in vehicular networks," IEEE Transactions on Vehicular Technology, vol. 58, no. 1, pp. 420–431, 2009.

[17] A. Agarwal, D. Starobinski, and T. D. Little, "Phase transition of message propagation speed in delay-tolerant vehicular networks," IEEE Transactions on Intelligent Transportation Systems, vol. 13, no. 1, pp. 249–263, 2012.

[18] E. Baccelli, P. Jacquet, B. Mans, and G. Rodolakis, "Highway vehicular delay tolerant networks: Information propagation speed properties," IEEE Transactions on Information Theory, vol. 58, no. 3, pp. 1743–1756, 2012.

[19] A. Kesting, M. Treiber, and D. Helbing, "Connectivity statistics of store-and-forward intervehicle communication," IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 1, pp. 172–181, 2010.

[20] K. Yin, X. B. Wang, and Y. Zhang, "Vehicle-to-vehicle connectivity on two parallel roadways with a general headway distribution," Transportation Research Part C: Emerging Technologies, vol. 29, pp. 84–96, 2013.

[21] Z. Zhang, G. Mao, and B. D. Anderson, "Stochastic characterization of information propagation process in vehicular ad hoc networks," IEEE Transactions on Intelligent Transportation Systems, vol. 15, no. 1, pp. 122–135, 2014.

[22] J. He, L. Cai, J. Pan, and P. Cheng, "Delay analysis and routing for two-dimensional VANETs using carry-and-forward mechanism," IEEE Transactions on Mobile Computing, vol. 16, no. 7, pp. 1830–1841, 2017.

[23] Y. H. Kim, S. Peeta, and X. He, "An analytical model to characterize the spatiotemporal propagation of information under vehicle-to-vehicle communications," IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 1, pp. 3–12, 2016.

[24] [dataset] U.S. Department of Transportation. (2017) Next Generation Simulation (NGSIM) vehicle trajectories and supporting data. (Accessed June 10, 2018). [Online]. Available: https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajector/8ect-6jqj

[25] C. Sommer, D. Eckhoff, A. Brummer, D. S. Buse, F. Hagenauer, S. Joerer, and M. Segata, "Veins: The open source vehicular network simulation framework," in Recent advances in network simulation. Springer, 2019, pp. 215–252.

[26] S. Eichler, "Performance evaluation of the ieee 802.11 p wave communication standard," in 2007 IEEE 66th Vehicular Technology Conference. IEEE, 2007, pp. 2199–2203.

[27] R. Besenczi, N. Bátfai, P. Jeszenszky, R. Major, F. Monori, and M. Ispány, "Large-scale simulation of traffic flow using markov model," Plos one, vol. 16, no. 2, p. e0246062, 2021.

[28] M. Ramezani and N. Geroliminis, "On the estimation of arterial route travel time distribution with Markov chains," Transportation Research Part B: Methodological, vol. 46, no. 10, pp. 1576–1590, 2012.

[29] J. Dong and H. S. Mahmassani, "Flow breakdown and travel time reliability," Transportation Research Record, vol. 2124, no. 1, pp. 203–212, 2009.

[30] N. Geroliminis and A. Skabardonis, "Prediction of arrival profiles and queue lengths along signalized arterials by using a Markov decision process," Transportation Research Record: Journal of the Transportation Research Board, vol. 1934, pp. 116–124, jan 2005.

[31] A. S. Alfa and M. F. Neuts, "Modelling vehicular traffic using the discrete time Markovian arrival process," Transportation Science, vol. 29, no. 2, pp. 109–117, 1995.