On Modeling and Understanding Vehicle Evacuation Attacks in VANETs

Mingkui Wei
Computer Science
Sam Houston State University
Huntsville, TX 77341

Zhuo Lu
Electrical Engineering
University of South Florida
Tampa, FL 33620

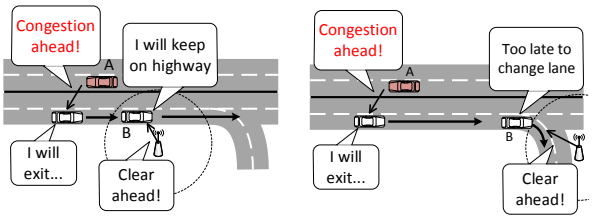
Wenye Wang
Electrical and Computer Engineering
North Carolina State University
Raleigh, NC, 27606

***Abstract*—To secure a Vehicular Ad-hoc Network (VANET), extensive studies have been conducted on developing authentication infrastructures, and identifying misbehaving vehicles. The effectiveness of such efforts heavily depends on the underlying communication network. However, information exchange in the VANET can be severely delayed because of its highly-dynamic and partially-connected topology. Such delay can be potentially exploited by attackers to cause physical impacts to the transportation system. In this paper, we propose and model a new attack, called *vehicle evacuation attack*, to investigate how the message delay endangers the trustworthiness in VANETs, and further causes physical impacts to cars on the road. Our study demonstrates that there exists a linear relationship between the delay of message dissemination and the impact of the vehicle evacuation attack, which can be used as a guideline on security, reliability, and safety design in real-world VANETs.**

I. Introduction

Facing the shear contrast between daily increasing volume of vehicles and the relatively under-developed highway system, people begin to exploit the potentials of the Intelligent Transportation Systems (ITS) for safer and more efficient human and freight transportation. An essential enabling feature of the ITS is the Vehicular Ad-hoc Network (VANET), which is established with the On-board Unites (OBUs), i.e., microcomputers equipped on vehicles to facilitate the Vehicle-to-Vehicle (V2V) communication. Over the V2V network, critical information such as road conditions or emergencies can be exchanged [1], which can result in enhanced transportation efficiency and reliability. While the V2V network is dedicated for inter-vehicle communication without Internet access, the Vehicle-to-Infrastructure (V2I) network enables vehicles to access Internet with the help of Road Side Units (RSUs).

The VANET is a highly-dynamic network because vehicles travel fast and cause the topology of the network to change rapidly. Stemming from this character, information trustworthiness becomes a major concern in VANETs [2]. For instance, two cars may have never met before, how can they know they should trust each other during V2V information exchange? And if a car is identified as malicious, how to inform other cars of its malicious identity? In this regard, many works have focused on identifying misbehavior in VANETs, or developing secure infrastructures to regulate vehicles' behavior. In particular, the work in [3] proposed a comprehensive secure architecture leveraging many modern security strategies. This architecture has its basis as a centralized Public Key Infrastructure (PKI), on which other strategies such as Intrusion Detection and Pseudonymity are further deployed.



(a) Malicious messages misled vehicles. (b) Malicious messages unable to mislead vehicles.

Fig. 1. Scenarios where VANET (a) is or (b) is not susceptible to malicious attacks.

The work in [4] improved the conventional PKI to a group signature design, which allows a group of vehicles to sign for a message such that individual vehicle's identity can be preserved. From the aspect of directly identifying misbehaving vehicles, [5] proposed to detect misbehavior nodes based on theirs deviations from an established model of the network, and [6] proposed another approach by observing both disseminated messages and the subsequent behavior of a particular vehicle. However, an assumption for these approaches to work is reliable real-time information exchange supported by communication networks in VANETs, which is, unfortunately, not always true in real-world scenarios.

Indeed, message exchange in VANETs can endure long delay and also is error-prone. In this paper, we find that such long message delay and partially-disconnected topology in a VANET result in a security vulnerability that can cause physical impacts on vehicles. In particular, we use examples in Fig. 1 to explain the potential security issue.

In Fig. 1 we present a highway segment with two directional roads. Vehicles travel from right to left (westbound) on the top lanes, and travel from left to right (eastbound) on the bottom lanes. As a malicious vehicle A travels westbound, it keeps broadcasting false information to eastbound vehicles, telling them that there is a traffic condition in front of them. In both Fig. 1(a) and Fig. 1(b), the benign vehicle B encounters A at a place outside the coverage of any RSU that provides Internet access. This means that B receives A's false information, but will experience a delay period before it can encounter an RSU and verify the information from A via the RSU. In Fig. 1(a), there is an RSU in the close proximity, thus B has the opportunity to reconnect and verify with the RSU the real traffic condition and hence discard the false information from A. However, as shown in Fig. 1(b), the delayed information

exchange between B and the RSU misleads B to exit the road to avoid the (actually non-existed) traffic. We call such an attack *vehicle evacuation attack*. Note that in this case, unless B already knows that A is malicious, which still remains as an open research issue [2], the impact of A on B is unavoidable.

As demonstrated by Fig. 1, the major cause for B to be vulnerable in Fig. 1(b) is the long communication delay between the time when B receives false information and the time when B can verify it with an RSU. While this issue can be easily addressed if we have enough RSUs to provide Internet access to all vehicles, this solution is infeasible in practice due to high RSU deployment costs [7]. Therefore, to evaluate the performance of such an ITS, it is no longer sufficient to consider only in the communication or network domain. For instance, no matter how much the delay will be, vehicle A can cause no harm to B if there is no Exit on the road (because the physical condition makes B have no choice). This observation motivates us to explore the research question: *what is the role of message delay in escalating the physical impacts of vehicle evacuation attacks in VANET*?

To answer this question, we first model the *vehicle evacuation attack*. In such an attack, a malicious vehicle keeps broadcasting false traffic information and tries to evacuate vehicles from the highway, as demonstrated in Fig. 1. The model is developed based on real-world traffic data, and is validated with intensive simulations. We define the metric *number of evacuated vehicles* (i.e., the total number of vehicles that are misled by false information and then exit) to evaluate the consequences of such attacks, and demonstrate that there exists a linear relationship between the worst case number of evacuated vehicles and message delay in a VANET. Moreover, the modeling of the *vehicle evacuation attack* also provides accurate statistical properties of vehicles and clusters of vehicles (i.e., a set of vehicles that are able to communicate with each other via single-hop or multi-hop communications) on the highway, which can be utilized by broader research regarding ITS and facilitate future studies. To our best knowledge, this is the first work considering the security of ITS jointly from the communication network domain and the physical domain.

The rest of this paper is organized as follows. In Section II, we mathematically model the *vehicle evacuation attack* based on the vehicle and cluster properties derived from real-world highway traffic data. In Section III, we explore the correlation between message delay and the consequence of this attack, and discuss potential solutions to combat this attack. In Section IV we conclude our work.

II. Vehicle Evacuation Attack on the Highway

In this section, we model the vehicle evacuation attack based on real-world highway traffic data.

A. Scenario and Attack Model

Consider the scenario shown in Fig. 2, where two consecutive RSUs are placed between a *segment* of a highway. Without loss of generality, we consider the location of the left-end RSU in Fig. 2 as the origin, and denote the length of this segment between these two RSUs as D.

At the beginning, we assume there is a malicious vehicle (i.e., an attacker) traveling westbound. Along the way, it keeps broadcasting false traffic information (i.e., there is a congestion) to eastbound vehicles, and these vehicles will react to the false information differently depending on the positions where they encounter the malicious vehicle. The eastbound vehicles in Fig. 2 can be classified into three categories.

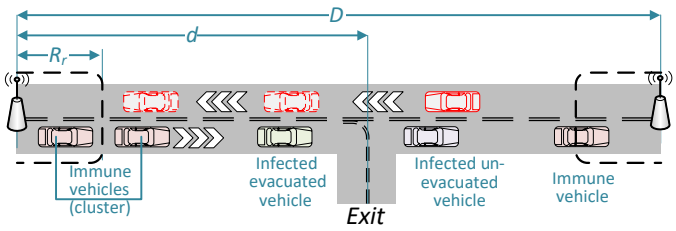


Fig. 2. Attack scenario with infected/evacuated/immune vehicles.

The *immune vehicles* are vehicles that are connected to at least one RSU when they encounter the attacker, either directly covered by an RSU, or indirectly bridged by a cluster of vehicles.

The *infected vehicles* are vehicles that are not connected to any RSU at the time when they encounter the malicious vehicle. In this case, the infected vehicle will choose to trust the false information, until it re-connects to an RSU.

The *evacuated vehicles* are vehicles that are infected, and there is an Exit between where it was infected and the next RSU. These vehicles will exit the highway at that Exit.

It is also easy to understand that, while multiple Exits may exist on this segment in Fig. 2, it is the location of the *last Exit* on the segment that determines the eventual number of evacuated vehicles, because after the Exit vehicles cannot exit as they can verify the information when reaching the next RSU. We denote the distance from the last Exit to the origin as d.

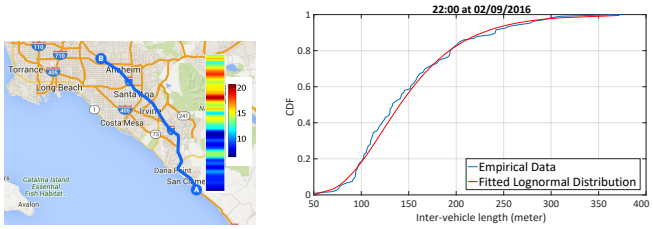
Notice that in practice a vehicle may not exit even if it receives false information. Thus, it is more practical to assume such a car has a certain probability to exit. In our modeling, we assume this probability to be 1 for simplicity and to show the maximum impact. We note that the value of the probability can be adjusted with straightforward extension in the formulation.

B. Distribution of Vehicles on the Highway

To model the behavior and interaction of vehicles and depict the consequence of a vehicle evacuation attack, it is critical to identify how vehicles are distributed on the road. In this subsection we adopt a data analysis approach to obtain the distribution characteristics of vehicles.

1) Obtaining the Dataset: Our model is developed based on analysis of real-world traffic data obtained from the California *Caltrans Performance Measurement System* (PeMS) [8]. The PeMS collects real-time data from more than 39,000 individual sensors placed on the freeway/highway system across all major metropolitan areas of the California State, and the data is then classified according to geographic highway sections. All data is managed and published by the California Department of Transportation.

In addition to common traffic-related data, the PeMS also keeps monitoring and reporting the health status of sensors



(a) I5-N District 12 in PeMS . (b) l_v at 22:00pm of Feb 9th, 2016.

Fig. 3. Freeway I5-N Location and Characters (colorbar indicates vehicle density, i.e., vehicles/mile).

at each highway section, since the sensors are not highly reliable and a fraction of them can malfunction at any time. The particular dataset that is used in this study is taken on Feb 9, 2016, and from the highway segment *I5-N District 12 (Orange County)*, because this segment provides the highest percentage of healthy sensors at the time of study. As a brief overview, the studied segment is the northbound of highway Interstate 5, which is 44.4 miles in length. It ranges from San Clemente to Buena Park, and passes the whole Orange County. The overall segment contains 333 sensors at the time of study. A snapshot of this highway segment is shown in Fig. 3(a).

2) Distance between Vehicles l_v: Since communication between vehicles is limited by their transmission range, it is critical to characterize the relative distance l_v between vehicles on the road. However, inter-vehicle distance is not directly measured by the PeMS system. We first make the following assumptions, and then demonstrate our procedure to infer statistical properties of l_v based on measured data.

1. The distance l_v measures the center-to-center distance of two consecutive vehicles. Because what we are interested in is the distance between the communication devices of two vehicles, without loss of generality, we can assume that such device is equipped at the center of each vehicle.

2. Vehicles between two consecutive sensors are evenly distributed, this is because the data granularity is limited by the density of sensors and we cannot obtain more detailed data between two sensors. However, our data shows the average distance between two sensors is about 210 meters, while even during the busiest hours, data samples of l_v are still larger than 50 meters. Therefore, statistically there are at most 4 vehicles between two sensors, thus this assumption will not cause significant impact to the accuracy of the result.

Based on above two assumptions, l_v can be calculated with the *occupancy*, which is a directly measured parameter at all sensors. The *occupancy* (denoted as θ) is a common parameter in transportation measurement, which is defined as the percentage of time over a given period of time (5 minutes in this study) that the detection zone is occupied by vehicles [8]. Based on θ, the inter-vehicle distance l_v can be calculated according to the linear equation:

$$l_v(meters) = \frac{\lambda(feet/vehicle)}{\theta \times 5280(feet/mile)} \times 1609.34(meters/mile), \tag{1}$$

in which λ denotes the average length of a vehicle.

To find the statistical properties of l_v, we choose 3 time instances during the day to study different traffic conditions, which are: 3:00am (light traffic), 7:00am (heavy traffic), and 22:00pm (medium traffic).

We assume the average vehicle length λ to be 20 feet ($\sim$6 meters) based on the report from the US Department of Transportation [9].

We apply curve-fitting in Matlab for l_v for 3 time instances, and observe that all of them can be fitted to the *lognormal* distribution with minimum errors. To further validate our observation, we perform the Chi-square goodness-of-fit test [10] between the empirical data and the fitted distribution. To begin with, we set the *null hypothesis* to be: the empirical data comes from the fitted lognormal distribution. As a result, the null hypothesis is accepted for all three data sets at significance level 0.05, which indicate these data follows the lognormal distribution very well. The visual comparison for the dataset at 22:00pm is provided in Fig. 3(b).

Based on this observation, in the following study we assume that the *inter-vehicle distance* l_v follows the *lognormal* distribution that is characterized by σ_v and μ_v, which are summarized in the following model.

Model 1: The probability density function (PDF) of the *inter-vehicle distance* l_v is given by:

$$f_{l_v}(x) = \frac{1}{x\sigma_v\sqrt{2\pi}} exp\left(-\frac{(\ln x - \mu_v)^2}{2\sigma_v^2}\right). \tag{2}$$

Accordingly, the expected value of l_v is:

$$E[l_v] = e^{\mu_v + \sigma_v^2/2}. \tag{3}$$

It is worth noting that our observation that l_v follows the lognormal distribution is supported by other recent studies as well. For instance, [1], [11] made the similar observation from a dataset that completely differs from ours.

3) Size and Length of Cluster of Vehicles: A cluster is a set of vehicles that are able to communicate with each other in either single-hop or multiple-hop manners. In the case that a vehicle is not directly covered by an RSU, it can still communicate with the RSU as long as it belongs to a cluster, and at least one vehicle in the cluster can communicate with the RSU. To this end, clusters play a critical role in facilitating normal operation of VANETs, and we aim to understand statistical properties of clusters. For easy demonstration, in the following we use R_v and R_r to denote the transmission ranges of a vehicle and an RSU, respectively.

Distribution of Cluster Size s_c: We use s_c to denote the size of a cluster, i.e., the number of vehicles in a cluster. When l_v follows the lognormal distribution, the cluster size s_c can be calculated by the following lemma.

Lemma 1: The cluster size s_c is a discrete random variable whose probability mass function (PMF) is given by:

$$Pr\{s_c = k\} = (1-p)^{k-1}p, \tag{4}$$

where $p = 1 - \Phi\left(\frac{\ln R_v - \mu_v}{\sigma_v}\right)$, $\Phi(x) = \frac{1}{2} + \frac{1}{2}erf(\frac{x}{\sqrt{2}})$, and $erf(x) = \int_0^x \frac{2}{\sqrt{\pi}} e^{-z^2} dz$.

Proof: From statistical perspective, the size of a cluster can be interpreted as the number of trials before the value

of a lognormal-distributed random variable first exceeds the threshold value R_v, which obviously follows the Binomial distribution. ■

Distribution of Cluster Length l_c: The *Cluster Length* l_c denotes the distance between the first and the last vehicle of a cluster, and it is essentially a sum of a random number of random variables. We use the following lemmas to characterize its statistical properties.

Lemma 2: The PDF of the cluster length l_c is given by:

$$f_{l_c}(x) = \sum_{k=1}^{\infty} Pr\{s_c = k\} \cdot f_{l_{c|k}}(x|s_c = k), \tag{5}$$

where $l_{c|k}$ is the length of a cluster with size $s_c = k$.

Proof: Consider a cluster composed by $k+1$ vehicles, let the cluster length be represented by $l_c = l_{v_1} + l_{v_2} + ... + l_{v_k}$, in which l_{v_i}, for $i \in [1, k]$, is the i^{th} inter-vehicle distance in this cluster. For a group of vehicles to formulate a cluster, it is required that $l_{v_i} \leq R_v,\ \forall i$, otherwise, it will break into more than one clusters. This requires l_{v_i} to be a truncated lognormal random variable with upper limit R_v, and we use $\bar{l_v}$ to denote this truncated lognormal random variable. Particularly,

$$\bar{l_v} \equiv \begin{cases} l_v & l_v \leq R_v \\ 0 & otherwise. \end{cases} \tag{6}$$

Then, applying the *law of total probability*, we obtain the PDF. ■

Although the expression in equation (5) looks simple and intuitive, it is in fact challenging to find the closed-form expression of $f_{l_c}(x)$. This is because $l_{c|k}$ is a sum of k truncated lognormal random variables, and a closed form expression of its PDF is not yet known [12]. We address this challenge by proposing an analytical approximation as follows.

Lemma 3: The PDF of the cluster length l_c given in Lemma 2 can be approximated as:

$$f_{l_c}(x) = \frac{(1-p)p}{x\sigma_{c|2}\sqrt{2\pi}} exp\left(-\frac{(\ln x - \mu_{c|2})^2}{2\sigma_{c|2}^2}\right)\chi_{R_v}(x) + \sum_{k=3}^{\infty} \frac{(1-p)^{k-1}p}{\sigma_{c|k}\sqrt{2\pi}} exp\left(-\frac{(x-\mu_{c|k})^2}{2\sigma_{c|k}^2}\right)\chi_{(k-1)R_v}(x), \tag{7}$$

where $\mu_{c|k} = (k-1)\cdot E[\bar{l_v}]$ and $\sigma_{c|k} = \sqrt{(k-1)\cdot(E[\bar{l_v}^2] - E[\bar{l_v}]^2)}$ for $k \geq 2$.

And $\chi_A(x) = \begin{cases} 1 & x \leq A, \\ 0 & otherwise, \end{cases}$ is a step function.

Proof: See Appendix. ■

C. Number of Evacuated Vehicles

In the following we demonstrate the derivation of the number of evacuated vehicles $N(d)$. It is assumed that the westbound and eastbound vehicles travel with approximately constant speed, which is mostly the case on a highway without any congestion, and the speeds are denoted as v_w and v_e, respectively. These denotations are also demonstrated in Fig. 2. The following theorem characterizes the number of evacuated vehicles.

Theorem 4: Given a segment with length D, on which the eastbound vehicles travels at speed v_e and with their inter-vehicle distance $l_v \sim lognormal(\mu_v, \sigma_v)$, and the westbound attacker travels at speed v_w, and starts to broadcast false information as it travels from distance D to the origin. Denote the distance of the last Exit to origin as $d \in [0, D]$. The number of evacuated vehicles is given by:

$$N(d) = \frac{(1+\nu)}{E[l_v]} \cdot \int_0^d \left(\int_0^{x-R_r} f_{l_c}(y)dy \cdot \int_0^{D-R_r-x} f_{l_c}(y)dy\right)dx, \tag{8}$$

where $\nu = v_e/v_w$ is the speed ratio of the eastbound to the westbound traffic.

Proof: The number of evacuated vehicles can be calculated by integrating the number of infected vehicles at each point of the segment from 0 to d.

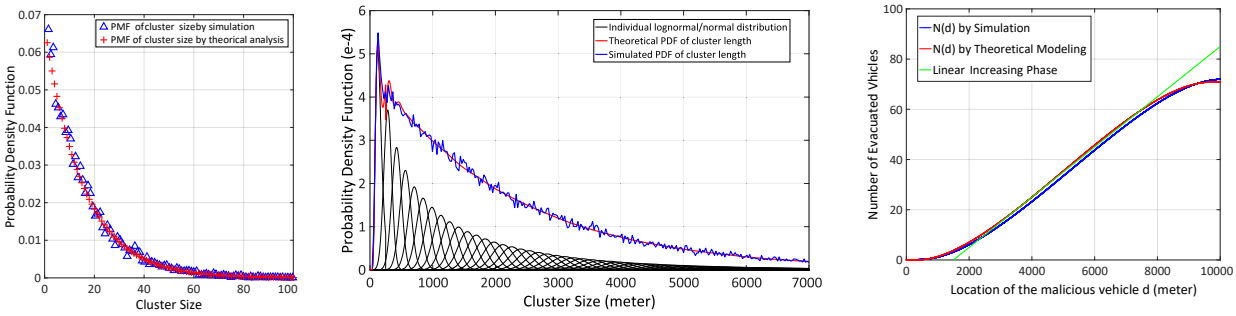
Denote an arbitrary point on the segment as x. If $x \in [0, R_r]$ or $x \in [D - R_r, D]$, that is, if this point is covered by RSUs, then any vehicle at this location has probability 0 to be infected because it can communicate directly with an RSU. On the other hand, for $x \in (R_r, D - R_r)$, the number of infected vehicles is the product of the two parameters: the probability a vehicle will be infected, and the average number of vehicles appear at $x \in (R_r, D - R_r)$.

The probability that a vehicle will be infected at at $x \in (R_r, D - R_r)$ is the probability that it is not connected to both RSUs, which can be calculated as $Pr\{vehicle\ \ disconnected\ \ at\ \ x\} = \int_0^{x-R_r} f_{l_c}(y)dy \cdot \int_0^{D-R_r-x} f_{l_c}(y)dy$.

In particular, we consider the interval $[x, x + \Delta x]$. Then, the number of vehicles appearing in this interval is the product of the vehicle density and the length of this interval. Because vehicles in the two lanes move towards opposite directions with different speeds, we have to choose one direction as the reference. We here choose to consider the problem from the malicious vehicle's perspective. The time it takes for the malicious vehicle to travel from x to 0 is x/v_w, during this time, the distance that an eastbound vehicle has traveled is $(x/v_w) \cdot v_e = \nu x$. Therefore, during this process, the total number of eastbound vehicles that actually encountered the malicious vehicle is $(x + \nu x)/E[L_v]$. On the other hand, the malicious vehicle has traveled only for a distance x, therefore, the density of the eastbound traffic, from the malicious vehicle's perspective, is $\frac{(1+\nu)x}{E[l_v]\cdot x} = \frac{1+\nu}{E[l_v]}$, and the number of vehicles between the interval $[x, x + \Delta x]$ is therefore $\frac{(1+\nu)}{E[l_v]} \cdot \Delta x$. Accordingly, the number of infected vehicles within this interval can be found as:

$$\frac{(1+\nu)\cdot\Delta x}{E[l_v]}\left(\int_0^{x-R_r} f_{l_c}(y)dy \cdot \int_0^{D-R_r-x} f_{l_c}(y)dy\right). \tag{9}$$

Taking integration of equation (9) from 0 to d, as $\Delta x \to 0$, completes the proof. ■



(a) PMF of Cluster Size. (b) PDF of Cluster Length. (c) $N(d)$ as d increases from 0 to D.

Fig. 4. Comparison between simulation and theoretical analysis.

Notice that the malicious vehicle may not be the only vehicle existing on the westbound lane, we assume other westbound vehicles do not help eastbound vehicles verify this information even they have access to RSUs, because this information regards to the situation behind of the westbound vehicles, and thus of no interests to them.

Remark 1: It is clear that $N(d)$ is an increasing function of $d \in [0, D]$, Therefore, $d = D$ (i.e., the last Exit has distance D to the origin) represents the worst case scenario where the number of infected vehicles is equal to the number of evaluated vehicles. Thus, we can use $N(D)$ to denote the *number of infected vehicles* in the worst case.

D. Validation

We validate the derived statistical properties by comparing the theoretical results to Matlab simulation results. We implement a straight two-lane highway segment, with one lane at each direction. The length of this segment is set to be 10,000 meters. At the beginning of each simulation, vehicles are generated on both lanes, and the inter-vehicle distance is set to be lognormally distributed with parameters obtained from the PeMS dataset at the time 22:00pm. The transmission range of both vehicles and RSUs is set to be 250 meters, which has been shown to be a practical value in existing works [1], [11].

1) Size and Length of Clusters: From the distribution fitting based on the 22:00pm data, we find the parameters for l_v are $\mu_v = 4.9468$, and $\sigma_v = 0.3747$. Accordingly, we solve that $\Phi\left(\frac{\ln R_v - \mu_v}{\sigma_v}\right) = 0.9374$.

Fig. 4(a) shows the PMF of s_c for both theoretical and simulation results, and Fig. 4(b) shows the PDF of l_c derived from both theoretical analysis and simulations. For the theoretical approximation, we set k ranging from 2 to 100, i.e., we neglect the probability that a cluster has 100 vehicles or more. As shown by these figures, our analytic results match the simulation results very well.

2) Number of Evacuated Vehicles: Fig. 4(c) shows the comparison between simulation and theoretical results of the number of evacuated vehicles $N(d)$ as d goes from 0 to D with $\nu = 1$. From the figure, we can observe that our theoretical model matches the simulation results. Furthermore, we observe that $N(d)$ dose not increase linearly as d grows from 0 to D. Instead, $N(d)$ grows slower where it is close to either RSUs. This phenomena that $N(d)$ exhibits provide effective suggestions on RSU placement to enhance highway security, which are discussed in detail in the next section.

III. Understanding the Vehicle Evacuating Attacks

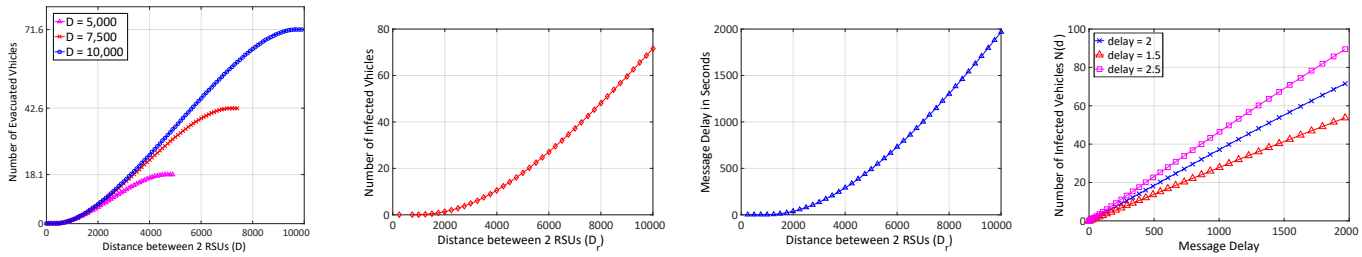
A. Understanding the Impacting Factors in $N(d)$

Recall that the number of infected vehicles $N(d)$ is defined by equation (8). Among all variables, R_r depends on specific technology and cannot be changed. In addition, $E[l_v]$ and $f_{l_c}(x)$ are determined by all drivers and are less likely to be manipulated. Therefore, $N(d)$ are essentially determined by three parameters: the relative speed ν, the location of the last Exit d, and the segment length D.

1) Impact of Relative Speed ν: From equation (8), we observe that the relative speed ν only appears in the linear part of this equation. Therefore, the change of ν will only change the significance of $N(d)$. While for most highways the speed limit are symmetrically set for both directions, the impact of ν can be escalated when the actual speed on opposite lanes experience significant difference, which is not uncommon in practice. For instance, in the metropolitan area where most people work in the downtown but live in satellite cities, traffic on both directions during rush hours can be significantly asymmetric. If one lane is experiencing a severe traffic jam while the opposite lane has light traffic, the malicious vehicle may choose to broadcast false information out of disgruntlement, or just for fun.

2) Impact of Last Exit Location d: The location of the last Exit d is the key factor that allows us to understand the physical impact of a malicious attack. While most existing works focus only on improving the cyber aspect of the ITS, our model demonstrates that the physical characteristics of the highway are equally, if not more, critical.

Fig. 4(c) clearly shows three stages of how $N(d)$ increases with d: super-linear from 0m to 3,000m, linear from 3,000m to 7,500m, and sub-linear from 7,500m to 10,000m. This knowledge provides useful suggestions on securing the highway in terms of RSU deployment: i) whenever possible, we want the last Exit to be located within the super-linear stage, where $N(d)$ grows slower than d, and ii) it is less favorable to have an Exit within the sub-linear stage because this provides less performance gain for the sake of reducing $N(d)$.



(a) $N(d)$ for three different value of D. (b) $N(D)$. (c) Message delay τ. (d) Relationship plot of $N(D)$ and τ.

Fig. 5. $N(D)$ and τ change as D increases from 250m to 10,000m with 250m each step.

3) Impact of Segment Length D: Fig.5(a) plots $N(d)$ for D to be 5,000m, 7,500m, and 10,000m. From Fig. 5(a), we observe that $N(D)$ (i.e., the worst-case $N(d)$) dose not appear to have a linear relationship with D. To better understand this observation, we plot in Fig. 5(b) the value of $N(D)$ as we increase D from 250m to 10,000m with a step of 250m.

From Fig. 5(b), we observe a super-linear relationship between $N(D)$ and D, which is consistent with our intuition: as D becomes larger, the benefit provided by a cluster gradually diminishes. When D is large enough such that the length-difference between a cluster and a single vehicle can be neglected, $N(D)$ will eventually become linear.

This observation suggests that in practice, we may want to choose the value of D to be smaller than the point where $N(D)$ becomes linear, because beyond this point the increase in D will incur larger negative impacts.

B. Correlating Message Delay and Vehicle Evacuation Attack

1) Per-segment Message Delay τ: We define the average message delay as the time interval from the time that a message is generated at a vehicle (i.e., when it wants to communicate with an RSU) to the time that this message is delivered at an RSU, which is essentially the average time an arbitrary vehicle on the segment encounters an RSU. We neglect any physical, transportation or application layer delay, since they are of several orders smaller than the delay we are interested in here. We provide the following result.

Theorem 5: For vehicles that travel with speed v, on a segment with length D, the average message delay, denoted as τ, satisfies

$$\tau = \int_0^{D-2R_r} f_{l_c}(x) \frac{(D-2R_r-x)^2(D-2R_r-2x)}{2vD^2} dx. \tag{10}$$

Proof: Without loss of generality, consider an eastbound vehicle that is in a cluster with length x. If this cluster's tail location is less than R_r, or its head location is greater than $D-R_r$, i.e., at least one end of this cluster is covered by one RSU, the delay for messages generated by this vehicle is $\tau_1 = 0$, and the probability for both cases is $(R_r+x)/D$.

For any other scenario, we can calculate the message delay as follows.

For this cluster to be isolated from any RSU, its tail location has to be larger than R_r, and head location has to be less than $D-R_r$. In this case, each vehicle in this cluster has equal probability to appear in an interval with length $D-2R_r-x$, thus, the probability that this cluster is not covered by any RSU is $(D-2R_r-x)/D$.

The message delay is proportional to the distance between the front RSU and the head of this cluster. Statistically, we can assume that the cluster is located at the middle of the segment, and the average travel time for a cluster to reach the front RSU is $(D-2R_r-x)/2v$.

The average message delay for a vehicle in a disconnected cluster with length x can therefore be calculated as:

$$\tau_2 = Pr\{l_c = x\} \cdot \frac{(D-2R_r-x)}{2v} \cdot \frac{(D-2R_r-x)}{D}. \tag{11}$$

And the probability for this scenario to happen is

$$Pr\{\tau_2\} = 1 - 2 \cdot \frac{R_r + x}{D}. \tag{12}$$

Moreover, a cluster is isolated implicitly means that $x < D-2R_r$. Then, we can calculate the average message delay as:

$$\tau = \tau_1 \frac{2R_r + 2x}{D} + \int_0^{D-2R_r} \tau_2 \cdot Pr\{\tau_2\} dx. \tag{13}$$

Bring equations (11) and (12) into (13) completes the proof. ■

The results in Theorem 5 are also validated and matched using simulations with the same setups in Section II-D. We omit such similar results due to the page limit.

2) Correlating $N(d)$ and τ: The average message delay τ is computed in equation (10), from which we are able to see that τ is determined by two factors: vehicle speed v and segment length D. As it is intuitive that τ linearly depends on vehicle speed v, we hereby only focus on understanding the impact of D. Figure Fig. 5(c) shows the change of τ as D increases from 250m to 10,000m with 250m each step, from which we observe that τ also has a super-linear relationship with D. Comparing Fig. 5(b) and Fig. 5(c), it is interesting to observe that these two figures are almost identical. And this observation indicates that $N(D)$ *is linearly related to* τ. To verify this, we plot the relationship between $N(D)$ and τ for three different speed ratio ν in Fig. 5(d), from which we are able to more clearly observe the linear relationship between these two parameters. Therefore, we can conclude that while the number of evacuated vehicles $N(d)$ can vary depending

on the location of the last Exit; its worst case scenario, $N(D)$, is linearly related to the average message delay.

3) Summary and Discussion: Recall that our objective is to identify the correlation between the average message delay and the consequence of the vehicle evacuation attack. In this paper, we achieve this goal by demonstrating that the message delay is linearly related to the worst case consequence of the *vehicle evacuation attack*. This observation enables us to more comprehensively inspect the characteristics of VANET and ITS systems by coupling the communication performance with the security performance, and provides insights in understanding the physical impact of such a malicious attack. Further security related studies can be conducted based on our observations. For instance, existing works on RSU placement consider this problem purely from economic point of view, i.e., how to cover more vehicles with less RSUs [7]. This work reminds us that road structures and their relative positions to RSUs also play critical roles in determining the security and reliability of the ITS system and should not be neglected.

IV. Conclusion

In this paper, we studied a new type of attack, vehicle evacuation attack, in the ITS from a new perspective. While most existing works still limit the ITS security study within the cyber domain, we take one step further and consider both the cyber and the physical domains. We demonstrate that partially connected VANETs with incurred delayed messages result in untrustworthy information, which eventually impacts the physical perspective of public transportation. Our work bears the significance in that it reminds us to consider the ITS security from both the cyber and the physical perspectives.

References

[1] A. B. Reis, S. Sargento, F. Neves, and O. Tonguz, "Deploying roadside units in sparse vehicular networks: what really works and what does not," *IEEE TVT*, vol. 63, no. 6, pp. 2794–2806, 2014.

[2] F. Qu, Z. Wu, F. Wang, and W. Cho, "A security and privacy review of vanets," *ITS, IEEE Trans.*, vol. 16, no. 6, pp. 2985–2996, 2015.

[3] K. Plossl, T. Nowey, and C. Mletzko, "Towards a security architecture for vehicular ad hoc networks," in *ARES, IEEE Proc.* IEEE, 2006.

[4] A. Studer, E. Shi, F. Bai, and A. Perrig, "Tacking together efficient authentication, revocation, and privacy in vanets," in *IEEE SECON*. IEEE, 2009, pp. 1–9.

[5] P. Golle, D. Greene, and J. Staddon, "Detecting and correcting malicious data in vanets," in *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*. ACM, 2004, pp. 29–37.

[6] S. Ruj, M. A. Cavenaghi, Z. Huang, A. Nayak, and I. Stojmenovic, "On data-centric misbehavior detection in vanets," in *Vehicular technology conference (VTC Fall), 2011 IEEE*. IEEE, 2011, pp. 1–5.

[7] T.-J. Wu, W. Liao, and C.-J. Chang, "A cost-effective strategy for road-side unit placement in vehicular networks," *Communications, IEEE Trans.*, vol. 60, no. 8, pp. 2295–2303, 2012.

[8] Caltrans PeMS. [Online]. Available: http://pems.dot.ca.gov/

[9] Office of the assistant secretary for research and technology, "Freight facts and figures 2015," *Bureau of Transportation Statistics*, 2015.

[10] M. Zhao, Y. Li, and W. Wang, "Modeling and analytical study of link properties in multihop wireless networks," *Communications, IEEE Trans.*, vol. 60, no. 2, pp. 445–455, 2012.

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

[12] L. F. Fenton, "The sum of log-normal probability distributions in scatter transmission systems," *Communications Systems, IRE Trans.*, vol. 8, no. 1, pp. 57–67, 1960.

[13] N. B. Mehta, J. Wu, A. F. Molisch, and J. Zhang, "Approximating a sum of random variables with a lognormal," *Wireless Communications, IEEE Trans.*, vol. 6, no. 7, pp. 2690–2699, 2007.

Appendix

The distribution of the lognormal sum (sum of regular lognormal random variables) has been studied for years [12], [13], but a closed-form expression is still unavailable. On the other hand, it has been shown that the lognormal sum can be approximated by a single lognormal random variable [13]. A well-known and accurate method in estimating parameters of this new lognormal random variable is the Fenton-Wilkinson (F-W) method [12], which is to match the first and second central moments of the new lognormal random variable to the original lognormal sum. However, the F-W method cannot be directly applied to solve equation (5), since the random variables to be added are truncated in this problem.

Inspired by the F-W method, we propose a suitable approximation of the distribution of the truncated lognormal sum. We use extensive Monte-Carlo simulations to create samples regarding the truncated lognormal sum, and we observe that its distribution well follows the normal distribution. To validate this observation, we perform the Chi-square goodness-of-fit. For each particular k, we generate 10,000 samples of the truncated lognormal sum and fit it to a normal distribution, and then test the goodness of fit between the sample data and the fitted distribution. We extensively simulated for $k \in [2, 500]$, and test their goodness-of-fit with significance levels 0.05 and 0.01. As a result, for significance level 0.05, there are 24 out of the 499 (4.8%) cases where the null hypothesis (i.e., sampled data matches the fitted distribution) is rejected, while for significance level 0.01, the number is only 7 (1.4%). The first rejected case is at $k = 9$, which contributes only little error considering $Pr\{s_c = 9\}$ is very small. Based on these results, we accept this approach as a good approximation for the truncated lognormal sum problem.

To estimate the parameters of this normal random variable, we adopt the principle of the F-W method. In particular, denote $l_{c|k}$ the length of a cluster with size k, we assume

$$\begin{cases} E[l_{c|k}] &= k \cdot E[\bar{l_v}], \\ Var(l_{c|k}) &= k \cdot Var(\bar{l_v}), \end{cases} \tag{14}$$

in which $E[\bar{l_v}]$ and $Var(\bar{l_v})$ can be easily derived by knowing the first and second moment of $\bar{l_v}$:

$$E[\bar{l_v}] = e^{\mu_v + \frac{\sigma_v^2}{2}} \cdot \frac{1 - \Phi(\sigma_v - \frac{\ln R_v - \mu_v}{\sigma_v})}{\Phi(\frac{\ln R_v - \mu_v}{\sigma_v})}, \tag{15}$$

and

$$E[\bar{l_v}^2] = e^{2\mu_v + 2\sigma_v^2} \cdot \frac{1 - \Phi(2\sigma_v - \frac{\ln R_v - \mu_v}{\sigma_v})}{\Phi(\frac{\ln R_v - \mu_v}{\sigma_v})}. \tag{16}$$

Summarizing equations (14), (15) and (16), the PDF of cluster length l_c in equation (5) is approximated as the result in Lemma 3 (recall that $l_c = 0$ when $k = 1$, and l_c is exactly the truncated lognormal random variable when $k = 2$).