Secure Message Dissemination in Vehicular Networks: A Topological Approach

Jieqiong Chen* and Guoqiang Mao*
*School of Electrical and Data Engineering, University of Technology Sydney, NSW 2007, Australia
Email: jieqiong.chen@student.uts.edu.au, g.mao@ieee.org

***Abstract*—Secure message dissemination is an important issue in vehicular networks, especially considering the vulnerability of vehicle to vehicle (V2V) message dissemination to malicious attacks. Traditional security mechanisms, largely based on message encryption and key management, can only guarantee secure message exchanges between known source and destination pairs. For secure message dissemination in vehicular networks against insider attackers, who may tamper the content of the disseminated messages, ensuring the consistency and integrity of the transmitted messages becomes a major concern that traditional message encryption and key management based approaches fall short to address. In this paper, by incorporating the underlying network topology information, we propose a novel heuristic decision algorithm that enables a vehicle to make a decision on the message content using minimal information readily available. The proposed algorithm can be readily implemented in practice. Simulations are conducted to compare the security performance achieved by the proposed decision algorithm with that achieved by existing ones that do not consider or only partially consider the topological information, to establish the effectiveness of the proposed algorithm. Our results show that by incorporating the network topology information, the security performance can be significantly improved. This work sheds light on the optimum algorithm design for secure message dissemination in vehicular networks.**

***Index Terms*—Vehicular networks, security, message dissemination, decision algorithm.**

I. Introduction

Considering the vulnerability of vehicle to vehicle (V2V) communications, message dissemination in vehicular networks is susceptible to malicious insider attacks, e.g., malicious vehicles who may spread false messages, tamper or drop the received messages to disrupt delivery of authentic messages. These attacks in vehicular networks could potentially result in catastrophic consequences like traffic congestion, traffic crash, even loss of lives, and therefore are significant security threats to transportation systems that must be thoroughly investigated before vehicular networks can be deployed.

Conventional security mechanisms, largely based on message encryption and key management [1], [2], are effective to guarantee message integrity against outsider attackers, however fall short of protecting the integrity of disseminated messages when there exist insider attackers who possess valid certificates that can pass the authentication process conducted by the certification authorities [3].

To keep the network message dissemination secure against insider attackers, the trustworthiness of each vehicle and the integrity of their transmitted messages are of great importance. Different from traditional security settings, in vehicular networks, information collection and dissemination are conducted by distributed vehicles. Quite often, information may be generated by or received from a vehicle that has never been encountered before. Moreover, the associated vehicular network topology is constantly changing considering that both V2V and vehicle to infrastructure (V2I) connections may emerge opportunistically. These unique characteristics may render the entity-based trust establishment approach [4] futile in vehicular networks because it is challenging to maintain a stable reputation value for unknown and fast-moving vehicles. Furthermore, safety-related vehicular network applications usually require vehicles to respond quickly to the received messages [5]. In such cases, determining the integrity of the received messages is of greater importance than the malicious vehicle detection. Therefore, decision algorithms based on data consistency and integrity check emerge, e.g., [6]–[8]. However, when a vehicle receives conflicting messages from different nearby vehicles, it is not straightforward to assess which message is true if focusing on data only while ignoring the underlying network topology information that tells where these messages come from. Indeed, messages coming from different paths can be correlated when these paths share some common nodes. For instance, multiple false messages may result from the same malicious vehicle shared by multiple paths. Therefore, taking the underlying topological information into consideration is essential and beneficial when designing decision algorithms for vehicles to conduct data consistency and integrity check.

In this paper, we consider vehicular networks containing insider malicious vehicles that may tamper the content of messages to disrupt their successful delivery. We are interested in investigating a topology-based decision algorithm to keep vehicles from being misguided by false messages. To the best of our knowledge, this is the first work that takes the underlying topology information into consideration when checking the consistency of messages for secure message dissemination. The novelty and major contributions of this paper are summarized as follows:

1) By utilizing the underlying network topology information, we propose a heuristic decision algorithm to cope with the issue of message inconsistency caused by insider malicious vehicles in the network, so as to reduce their impact on the message security performance.
2) The proposed heuristic decision algorithm enables a

vehicle to make a decision using readily available information, so that is fairly easy to implement in practice.
3) Simulation results show that our proposed algorithm outperforms existing decision algorithms that do not consider or only partially consider the topological information in terms of secure message dissemination in vehicular networks.

The rest of this paper is organized as follows: Section II introduces the system model and the problem formation. Section III introduces the heuristic decision algorithm. In Section IV, we conduct simulations to validate its effectiveness and discuss its insight. Section V concludes this paper.

II. System Model and Problem Formation

A. Network and Message Dissemination Model

We consider a vehicular network where there is a vehicle (termed as *source vehicle*) intending to deliver a message about the road and traffic condition to inform other vehicles far away. The road condition information can be abnormal situations, e.g., traffic accident, slippery road, etc., or normal situation, e.g., uncongested traffic. We assume that the content of message takes value from $\{0, 1\}$, where 1 and 0 represents abnormal and normal road condition respectively. We denote the content of message transmitted by the source vehicle, which represents the actual road condition, by m_0, $m_0 \in \{0, 1\}$. Other vehicles do not know the true value of m_0 *a priori*.

The message is forwarded from the source vehicle in a broadcast and multi-hop manner to other vehicles with the help of relay vehicles. Relay vehicles can be any vehicle along the message propagation path. Multi-path forwarding makes it challenging for the attackers to influence all message forwarding paths [6], therefore helps to improve the message security. We consider each vehicle has a unique ID number that is registered in certification authority to represent its identity, and vehicles cannot forge their own or other vehicles' ID numbers. When a vehicle transmits a message to other vehicles, it adds its identity information, i.e., ID number, to the message. This is commonly adopted in vehicular network applications and can be achieved by some standard signature approach [9]. Using this, any vehicle in the network is able to obtain an integrity-protected path list of its received messages recording the relay vehicles of each message, and the records cannot be injected and removed by attackers.

B. Attack Model

We consider insider attackers in this paper, that is, we assume all the vehicles are legitimate vehicles that have passed the authentication process conducted by the certification authority [7]. Vehicles in the network can be classified into two categories: *normal vehicles,* which behave normally and will forward the received message without any alteration, and *malicious vehicles,* which may tamper the received message. Malicious vehicles are uniformly distributed among all the vehicles in the system with proportion p. It follows that the probability of each vehicle being a malicious vehicle is p, independent of the event that another distinct vehicle is a malicious vehicle. We initially assume that p is known and show that the accurate decision algorithm depends on the knowledge of p. However, in practice p is rather difficult to estimate and therefore we present a heuristic algorithm that removes the need to know p.

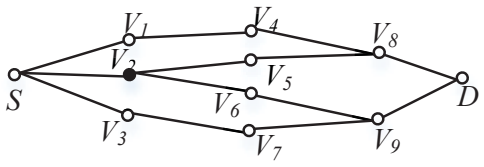

Fig. 1. An illustration of a vehicular network when there exists a malicious vehicles V_2 who would tamper the content of message. Specifically, there are four paths ($S-V_1-V_4-V_8-D$, $S-V_2-V_5-V_8-D$, $S-V_2-V_6-V_9-D$, and $S-V_3-V_7-V_9-D$) that deliver the source message from S to D. Therefore, out of the four copies of messages received by D, two copies are incorrect as there are two paths containing the malicious vehicle V_2.

Without loss of generality, we assume that the source vehicle is normal and only relay vehicles may be malicious. The normal vehicles do not know which vehicles are normal or malicious. On the contrary, malicious vehicles not only know which vehicle is malicious, but also are capable of communicating with each other via back channels of infinite bandwidth [10]. That is, we assume malicious vehicles know what the correct message transmitted by the source vehicle is. As a consequence, each malicious vehicle simply transmit the incorrect message, i.e., different from message m_0, to its neighbors. This implies that as long as a message is relayed by at least one malicious vehicle, the message would be incorrect. Fig. 1 gives a simple example of message dissemination process when there are insider attackers in the network.

C. Problem Formation

Now we give a detailed description of the research problem considered in this paper. We consider that there is a vehicle, which is several hops away from the source vehicle, trying to make a decision on the message content when it receives several messages, and we call it the *destination vehicle*. Note that the destination vehicle can be any vehicle along the message dissemination path. From the time instant the destination vehicle receives the first message, it waits time period T to receive more messages before making a decision. T characterizes the response time requirement on the decision, and a larger T potentially allows the vehicle to receive more messages. We will discuss its impact on the integrity of the decision later in the simulation. Let k be the number of messages received by the destination vehicle during its waiting time period T and let n be the number of relay vehicles that participate in relaying these k messages from the source vehicle to the destination vehicle. Consequently, the network being considered has n relay vehicles and k paths between source and destination, and the values of k and n can be readily obtained by the destination vehicle from the received messages. Each relay vehicle simply add their ID and re-broadcast the received messages, except the malicious relay vehicles who would broadcast the incorrect message.

Denote the k messages received by the destination vehicle by M_i, $i = 1, 2, ...k$, $M_i \in \{0, 1\}$. As each message corresponds

to a specific delivery path from the source vehicle to the destination vehicle, we number the corresponding paths by $L_1, L_2, ...L_k$. In addition, we number the relay vehicles by $V_1, V_2, ...V_n$. A vehicle V_i may belong to one or more paths. We construct a topology matrix to represent the underlying network topological correlation. Specifically, based on the path information derived from the received messages, the destination vehicle can readily construct a $k \times n$ topology matrix B, where each row represents a path, each column a node (vehicle), and the (i, j)-th entry B_{ij} being an indicator whether vehicle V_j belongs to path L_i:

$$B_{ij} = \begin{cases} 1, & \text{if vehicle } V_j \text{ belongs to path } L_i \\ 0, & \text{else} \end{cases} \tag{1}$$

In this paper, we are interested in investigating a decision algorithm for the destination vehicle to make a correct decision on the disseminated message content against attacks from malicious vehicles by utilizing the underlying network topology information. Denote by d, $d \in \{0, 1\}$ the decision made by the destination vehicle. If $d = m_0$, we say the destination vehicle makes a correct decision, otherwise we say it makes an incorrect decision. We use the probability of correct decision, denoted by P_{succ}, as the performance metric to measure the secure message dissemination performance, and P_{succ} can be formally defined as follows:

$$P_{succ} = \text{Pr}(d = 1, m_0 = 1) + \text{Pr}(d = 0, m_0 = 0) \tag{2}$$

III. Heuristic Decision Algorithm

In this section, we propose a heuristic decision algorithm for the destination vehicle to make a decision when receiving conflicting messages purely based on the network topology information.

According to the Maximum Likelihood Estimation [11], given the k messages $M_1 = m_1, ...M_k = m_k$ received by the destination vehicle, a decisions can be made according to the following rule:

$$d = \begin{cases} 1, & \frac{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=1)}{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=0)} > 1 \\ 0, & \frac{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=1)}{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=0)} < 1 \end{cases}, \tag{3}$$

and when $\frac{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=1)}{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=0)} = 1$, d is randomly chosen from 0 and 1 with equal probability.

It can be seen from (3) that, the decision on $d = 1$ or $d = 0$ is made by comparing the value of $\frac{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=1)}{\text{Pr}(M_1=m_1,...M_k=m_k|m_0=0)}$ and 1. Therefore, calculating the probability that the event $M_1 = m_1, ...M_k = m_k$ occurs if the true message m_0 is 1, denoted as $\text{Pr}(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and the probability that the event occurs if the true message m_0 is 0, denoted as $\text{Pr}(M_1 = m_1, ...M_k = m_k|m_0 = 0)$, is the critical part of making a decision. In the following, we will first demonstrate a method of re-arranging the topology matrix B, followed by the calculation of the above two probabilities, and finally we present the detailed heuristic decision algorithm. Without loss of generality, we assume that among the k copies of messages $M_1 = m_1, ...M_k = m_k$ received by the destination vehicle, there are exactly k_1, messages with content 1 and the other $k - k_1$ messages with content 0. Note that $k_1 = 0$ and $k_1 = k$ are both trivial cases implying no conflict in the received messages so that the decision is straightforward, therefore we only consider the case when $0 < k_1 < k$.

A. Re-arranging the topology matrix B

Based on the message content delivered by different paths, we re-arrange the network topology matrix B into the following form:

$$B = \begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \\ \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}, \tag{4}$$

where B_1, B_0, B_{s_1} and B_{s_0}, if exist, are non-zero matrices, and $\begin{bmatrix} B_1 & B_{s_1} & \mathbf{0} \end{bmatrix}$ is a $k_1 \times n$ sub-matrix corresponding to the paths that deliver messages with content 1 to the destination vehicle, and $\begin{bmatrix} \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$ is a $(k - k_1) \times n$ sub-matrix corresponding to the paths that deliver messages with content 0 to the destination vehicle. Besides, the columns of B_1 and B_0 correspond to vehicles that only belong to paths that deliver messages with content 1 and paths that deliver messages with content 0 to the destination vehicle respectively, denoted by *Type 1* vehicles and *Type 0* vehicles respectively. The columns of sub-matrix $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$ correspond to vehicles that belong to at least one path that delivers message with content 0 and one path that delivers message with content 1 to the destination vehicle, denoted by *Type 2* vehicles. Assume the number of Type 1 and Type 0 vehicles are n_1 and n_0 respectively, $0 \leq n_1 + n_0 \leq n$, and the number of Type 2 vehicles is $n_2 = n - n_1 - n_0$.

It is worth noting that the above arrangement of columns and rows of matrix B corresponds to a re-numbering of vehicles and paths and it does not change the underlying topology in terms of path information. Besides, the sub-matrix B_1 can be non-existent if $n_1 = 0$, i.e., when the paths that deliver messages 0 to the destination vehicle contains all the n vehicles in the network. Under this circumstance, $B = \begin{bmatrix} B_{s_1} & \mathbf{0} \\ B_{s_0} & B_0 \end{bmatrix}$. Similarly, the sub-matrix B_0 (or $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$) can also be non-existent or empty when $n_0 = 0$ (or $n_2 = 0$).

B. Calculation of $Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and $Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)$

In this part, we show the method of calculating the two conditional probabilities $\text{Pr}(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and $\text{Pr}(M_1 = m_1, ...M_k = m_k|m_0 = 0)$ based on the re-arranged topology matrix B. The following two theorems summarize the results.

Theorem 1. *Consider that a destination vehicle receives k copies of message $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, and among which k_1 messages are with content 1 and the other $k - k_1$ messages are with content 0, $0 < k_1 < k$. Conditioned on the source message $m_0 = 1$, the conditional probability of the occurrence of event $M_1 = m_1, ...M_k = m_k$ can be calculated as follows:*

$$Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1) = \begin{cases} (1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i (1-p)^{n_0-i}\right], & n_0 > 0 \\ 0, & n_0 = 0 \end{cases}, \tag{5}$$

where n_0 is the number of Type 0 vehicles in the network, and $a_i, i = 1, 2, ...n_0$ is the number of combinations with exactly i malicious Type 0 vehicles leading to the occurrence of event $M_1 = m_1, ...M_k = m_k$.

Proof: When $n_0 = 0$, there are no Type 0 vehicles in the network, and the topology matrix $B = \begin{bmatrix} B_1 & B_{s_1} \\ \mathbf{0} & B_{s_0} \end{bmatrix}$. Under this circumstance, conditioned on the source message $m_0 = 1$, when the event that k_1 messages are with content 1 occurs, all the n vehicles in the network should be normal vehicles. It follows that the event that the other $k - k_1$ messages are with content 0 occurs with probability 0. Therefore, we have $\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1) = 0$ when $n_0 = 0$.

When $n_0 > 0$, we can conclude that if the matrix $\begin{bmatrix} B_{s_1} \\ B_{s_0} \end{bmatrix}$ exists, then the corresponding Type 2 vehicles should be all normal vehicles. Observing that there is no possibility for two paths sharing the same malicious vehicle to deliver different contents. Therefore, malicious vehicles exist either among Type 1 vehicles or among Type 0 vehicles.

Given the source message $m_0 = 1$, all the Type 1 vehicles should be normal vehicles. Malicious vehicles can only exist among Type 0 vehicles. Besides, the malicious Type 0 vehicles should be able to compromise all the $k - k_1$ paths (corresponding to the sub-matrix $\begin{bmatrix} \mathbf{0} & B_{s_0} & B_0 \end{bmatrix}$) to cause the occurrence of the event that all the $k - k_1$ paths delivering messages with incorrect content 0. Therefore, any combination of malicious vehicles should satisfy the follows condition: by implementing element-wise *union* operation on their corresponding columns in sub-matrix B_0, i.e., implementing element-wise Boolean operation *OR* on them, the result should be a column with each entry be 1.

Note that the number of malicious type 0 vehicles can be any integer within $[1, n_0]$. Denote by event e_i that randomly choosing i columns from sub-matrix B_0 and then conducting element-wise union operation on them, there results a column with each entry being 1. Denote by $a_i, i = 1, ...n_0$ the total number of combinations that event e_i occurs. Thus,

$$a_i = \sum_{j=1}^{z_i} I\left(\text{event } e_i \text{ occurs}\right), \tag{6}$$

where $z_i = \begin{pmatrix} n_0 \\ i \end{pmatrix}$, and $I(x)$ is an indicator function that $I(x) = 1$, when x is true; otherwise $I(x) = 0$.

It then follows from the combination theory [12] that :

$$\begin{aligned} &\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1) \\ &= (1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i (1-p)^{n_0 - i} \right], \end{aligned} \tag{7}$$

where the first part corresponds to the probability that the k_1 paths deliver messages with correct content 1, so that all the $n - n_0$ vehicles contained in these k_1 paths are therefore normal vehicles; and the second part is the probability that the $k - k_1$ paths deliver messages with incorrect content 0, which summing up all the probabilities of different malicious vehicle combinations. ■

Theorem 2. *Consider that a destination vehicle receives k copies of message $M_1 = m_1, M_2 = m_2, ...M_k = m_k$, and among which k_1 messages are with content 1 and the other $k - k_1$ messages are with content 0, $0 < k_1 < k$. Conditioned on the source message $m_0 = 0$, the conditional probability of the occurrence of event $M_1 = m_1, ...M_k = m_k$ can be calculated as follows:*

$$\begin{aligned} &Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0) \\ &= \begin{cases} (1-p)^{n-n_1} \cdot \left[\sum_{i=1}^{n_1} b_i \cdot p^i (1-p)^{n_1 - i} \right], & n_1 > 0 \\ 0, & n_1 = 0 \end{cases}, \end{aligned} \tag{8}$$

where n_1 is the number of Type 1 vehicles in the network, and $b_i, i = 1, 2, ...n_1$ is the number of combinations that exactly i malicious Type 1 vehicles leading to the occurrence of event $M_1 = m_1, ...M_k = m_k$.

Denote by event $e_i^{'}$ that randomly choosing i columns from sub-matrix B_1 and then conducting element-wise union operation on them, there results a column with each entry being 1. Denote by $b_i, i = 1, 2, ...n_1$ the total number of combinations that event $e_i^{'}$ occurs. Then, we have

$$b_i = \sum_{j=1}^{z_i^{'}} I\left(\text{event } e_i^{'} \text{ occurs}\right), \tag{9}$$

where $z_i^{'} = \begin{pmatrix} n_1 \\ i \end{pmatrix}$. Therefore, this theorem can be readily proved following the same method as that used in the proof of Theorem 1, and hence is ignored.

C. Heuristic Decision Algorithm

By combining (3), (5) and (8), it is readily to have $d = \begin{cases} 0, & n_0 = 0 \\ 1, & n_1 = 0 \end{cases}$, and when $n_0 > 0$ and $n_1 > 0$,

$$\begin{aligned} &\frac{\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 1)}{\Pr(M_1 = m_1, ...M_k = m_k | m_0 = 0)} \\ =&\frac{(1-p)^{n-n_0} \cdot \left[\sum_{i=1}^{n_0} a_i \cdot p^i (1-p)^{n_0 - i} \right]}{(1-p)^{n-n_1} \cdot \left[\sum_{i=1}^{n_1} b_i \cdot p^i (1-p)^{n_1 - i} \right]} \\ =&\frac{\sum_{i=1}^{n_0} a_i \cdot \left(\frac{p}{1-p} \right)^i}{\sum_{i=1}^{n_1} b_i \cdot \left(\frac{p}{1-p} \right)^i}. \end{aligned} \tag{10}$$

Recall that Type 2 vehicles (if any) can not be malicious vehicles. Therefore, when considering potential malicious vehicle combinations, we only focus on Type 1 and Type 0 vehicles. Specifically, we regard the network corresponding to sub-matrix B_1 and B_0 as networks that each row represents a complete path and each column represent a vehicle, denoted by network T_1 and network T_0 respectively. In the following, with a twist use of the term vertex cut set [13] from graph theory which defines a vertex set whose removal would disconnect a graph, we define *malicious cut set*, *size* of a malicious cut set, and *minimal malicious cut set* of a network in this paper, and demonstrate that the parameter a_i, $1 \le i \le n_0$ and b_i,

$1 \leq i \leq n_1$ in (10), which was defined in (6) and (9), are exactly the number of malicious cut sets with size i in the networks T_0 and T_1 respectively.

Definition 3. A *malicious cut set* of a network is a combination of vehicles, where if all vehicles in the set are malicious vehicles, all paths of the network can be compromised. The *size* of a malicious cut set is the number of vehicles contained in the set. A *minimal malicious cut set* is a malicious cut set with the smallest size.

Based on Definition 3, if a vehicle set is a malicious cut set, then each path of the network contains at least one vehicle belonging to this set. Recall that a_i (or b_i) represents the number of combinations that randomly choosing i columns from sub-matrix B_0 (or B_1) and then conducting element-wise union on them, there results a column with each entry being 1. That is, a_i (or b_i) represents the number of combinations that by choosing i vehicles from Network T_0 (or T_1) to form a vehicle set, each path of network T_0 (or T_1) contains at least one vehicle belongs to this set. Therefore, a_i, $1 \leq i \leq n_0$ and b_i, $1 \leq i \leq n_1$ are exactly the number of malicious cut sets with size i of the network T_0 and T_1 respectively.

According to the properties of malicious cut sets, it can be readily obtained that $a_i = 0$ if $a_{i+1} = 0$, and $a_{i+1} > 0$, if $a_i > 0$. Similarly, we have $b_i = 0$ if $b_{i+1} = 0$, and $b_{i+1} > 0$, if $b_i > 0$. Define

$$r_0 = \min\{i:\ a_i > 0\},\ \ 1 \leq r_0 \leq n_0 \tag{11}$$

and

$$r_1 = \min\{i:\ b_i > 0\},\ \ 1 \leq r_1 \leq n_1, \tag{12}$$

the smallest integer that satisfies $a_i > 0$ and $b_i > 0$ respectively. Therefore, r_0 and r_1 are the sizes of the minimal malicious cut set of network T_0 and T_1 respectively, and a_{r_0} and b_{r_1} are the number of minimal malicious cut sets of network T_0 and T_1 respectively. It follows that

$$\begin{aligned}\frac{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)}{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)} =& \frac{\sum_{i=r_0}^{n_0} a_i \cdot \left(\frac{p}{1-p}\right)^i}{\sum_{i=r_1}^{n_1} b_i \cdot \left(\frac{p}{1-p}\right)^i} \\ \approx& \frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}},\end{aligned} \tag{13}$$

where the first step is obtained from the fact that $a_1 = a_2 = ... = a_{r_0-1} = 0$, $a_{r_0} > 0$, and $b_1 = b_2 = ... = b_{r_1-1} = 0$, $b_{r_1} > 0$, and the second step is obtained by only keeping the first item of both the numerator and denominator. Considering the fact that when p is small, the probability that there are $i+1$ malicious vehicle in the network is much smaller than the probability that there are i malicious vehicles in the network, therefore, this approximation is quite accurate.

Note that when p is small, we have $\frac{p}{1-p} \ll 1$. Therefore, when $r_0 \neq r_1$, whether the value of $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}$ is larger than 1 is dominantly determined by the value of $r_0 - r_1$. Specifically, when $r_0 < r_1$, we have $\left(\frac{p}{1-p}\right)^{r_0-r_1} \gg 1$. In this case, the coefficient $\frac{a_{r_0}}{b_{r_1}}$ plays a marginal role and therefore $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}} > 1$; when $r_0 > r_1$, we have $\left(\frac{p}{1-p}\right)^{r_0-r_1} \ll 1$, and therefore $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}} < 1$. On the contrary, when $r_0 = r_1$, whether the value of $\frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}}$ is larger than 1 would heavily depend on the value of the coefficient $\frac{a_{r_0}}{b_{r_1}}$. Consequently, we have

$$\begin{aligned}\frac{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)}{\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)} \approx& \frac{a_{r_0}\left(\frac{p}{1-p}\right)^{r_0}}{b_{r_1}\left(\frac{p}{1-p}\right)^{r_1}} \\ & \begin{cases} > 1, & r_0 < r_1 \\ < 1, & r_0 > r_1 \\ = \frac{a_{r_0}}{b_{r_1}}, & r_0 = r_1 \end{cases},\end{aligned} \tag{14}$$

which shows that to compare the values of $\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 1)$ and $\Pr(M_1 = m_1, ...M_k = m_k|m_0 = 0)$, we only need to compare the values of r_0 and r_1, namely, the sizes of minimal malicious cut sets of network T_0 and T_1 when $r_0 \neq r_1$, or the values of a_{r_0} and b_{r_1}, namely, the number of minimal malicious cut sets of networks T_0 and T_1 when they have the same size of minimal malicious cut set.

From Menger's Theorem [13], the size of the minimal vertex-cut is equal to the maximum number of vertex-independent paths between these two non-adjacent vertices. Therefore, it is readily to conclude that r_0 and r_1 are also the numbers of maximum number of node-disjoint paths in networks T_0 and T_1 respectively. Note that calculating the maximum number of vertex-disjoint paths from source to destination is a special case of finding the maximum flow problem by setting every vertex capacity to be 1 [13]. Therefore, the values of r_0 and r_1 can be readily obtained by existing maximum flow algorithms, e.g., those introduced in [13], [14]. When $r_0 = r_1$, a_{r_0} and b_{r_1} can be obtained by an exhaustive search algorithm according to their definitions given by (6) and (9).

By combining (3) and (14), the decision rule of our proposed heuristic algorithm can be shown as

$$d = \begin{cases} 1, & (r_0 < r_1)\ or\ (r_0 = r_1,\ a_{r_0} > b_{r_1}) \\ 0, & (r_0 > r_1)\ or\ (r_0 = r_1,\ a_{r_0} < b_{r_1}) \end{cases}, \tag{15}$$

and when $r_0 = r_1$, and $a_{r_0} = b_{r_1}$, d is randomly chosen from 0 and 1 with equal probability.

Remark 4. The implication of the heuristic decision algorithm (15) can also be explained straightforwardly as follows. Given two networks that deliver conflicting message contents, by removing the common nodes shared by these two networks and regarding each path after the removal of the common nodes as a completely new path, there results in two new independent networks that deliver conflicting message contents. Therefore, decision can be made by comparing the robustness of the two new networks. Note that a larger size of the minimal malicious cut set of a network implies a larger number of minimal malicious vehicles are required to compromise that network, and consequently, a lower probability to deliver incorrect messages. Therefore, the decision will always be

chosen as the message delivered by the network with a lower probability to be compromised.

From (15), we can see that the proposed heuristic decision algorithm is purely based on topological information so that is fairly easy to implement in practice. In summary, the heuristic decision algorithm works as detailed in Algorithm 1.

Algorithm 1 Heuristic Decision Algorithm

INPUT: $M_1 ... M_k$
OUTPUT: d
begin

1) Construct topology matrix B based on the paths information derived from the received k copies of message;
2) Based on the constructed topology matrix B, calculate r_0 and r_1 based on maximum flow algorithm;
3) **If** $r_0 < r_1$ **then** $d = 1$

   elseif $r_0 > r_1$ **then** $d = 0$

   else calculate a_{r_0} and b_{r_1} based on their definition given by (6) and (9);

   if $a_{r_0} > b_{r_1}$ **then** $d = 1$

   elseif $a_{r_0} < b_{r_1}$ **then** $d = 0$

   else d is randomly chosen from 0 and 1 with equal probability

   end

   end

end

IV. Simulation and Discussion

In this section, we conduct simulations to establish the validity and effectiveness of the decision algorithms proposed in Section III. We generate a network that vehicles are Poissonly distributed in the road with density ρ, and each relay vehicle has a probability p to be a malicious vehicle. Vehicles communicate with their neighbors adopting the unit disk model [15] with a transmission range $r_0 = 250$m [16]. Messages are disseminated in a broadcast and multi-hop manner, and the per-hop transmission delay is set to be $\beta = 4$ms [16]. We focus on a destination vehicle located at a distance L from the source vehicle. From the time instant the destination vehicle receives the first message that reports road condition, it waits time period T to receive more number of messages before it starts to make a decision. At each simulation, the destination vehicle makes a decision given the received messages and the derived underlying topology information utilizing the proposed heuristic decision algorithm. The decision result can be either correct or incorrect. The simulation is repeated 5000 times and the proportion of the correct decision, i.e., the probability of correct decision P_{succ}, is plotted.

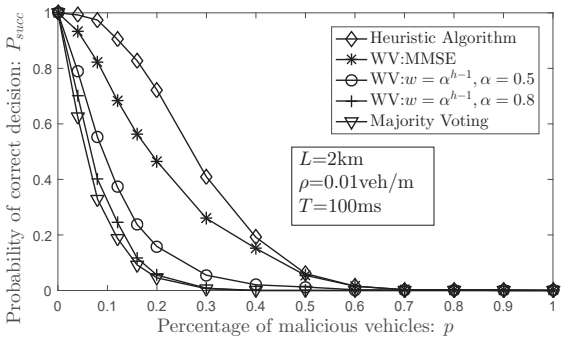

Fig. 2. A comparison of the probability of correct decision achieved assuming our proposed algorithm and that achieved assuming other existing weighted voting algorithms.

Fig. 2 compares the security performance achieved by our proposed heuristic algorithm (labeled: Heuristic Algorithm), with that achieved by existing weighted voting algorithms like the weighted voting algorithm proposed in [17] (labeled with WV: MMSE) that considers partial correlation between messages, the weighted voting algorithm proposed in [18] (labeled with WV: $w \propto \alpha^{h-1}$) that does not consider the underlying topology information causing the correlation between messages, and the majority voting (a special case of weighted voting by assigning identical weights to each vote) that totally ignores the underlying topological correlation. Specifically, the weighted voting algorithm proposed in [17] set weight to each message as $w_i = \sum_{j=1}^{k} C_{ij}^{(-1)} \left(\sum_{r,j=1}^{k} C_{rj}^{(-1)} \right)^{-1}$, where C is the error covariance matrix whose (i,j)th entry is defined by the error covariance between message M_i and message M_j, calculated by $C_{ij} = E\left[(M_i - m_0)(M_j - m_0)\right]$. C^{-1} is the inverse matrix of the error covariance matrix C, and $C_{ij}^{(-1)}$ is the (i,j)th entry of the matrix C^{-1}. The weighted voting algorithm proposed in [18] simply assigns weight to each message as $w_i = \frac{\alpha^{h_i - 1}}{\sum_j \alpha^{h_j - 1}}$, where $\alpha \in (0,1)$ is a weighting factor to reduce the oversampling impact caused by messages generated from the same source and h_i is the number of hops traveled by the ith message from the source to the destination. It can be seen that our proposed algorithm outperforms the weighted voting algorithms proposed in [17], [18], and the majority voting algorithm, which demonstrates that our algorithm that takes the topology information and correlation between different copies of message into account is able to effectively improve the robustness of vehicle networks against attacks from malicious vehicles.

Fig. 2 also reveals the relationship between the probability of correct decision P_{succ} and the percentage of malicious vehicles in the network. It can be seen that P_{succ} reduces to its minimum value $P_{succ} = 0$ when the proportion of malicious vehicles in the network is larger than a certain threshold. Beyond that threshold, a further increase in p has little impact on the security performance. This can be explained by the fact that the more malicious vehicles in the network, the more tampered copies of message will be delivered, and therefore a lower chance for the destination vehicle to make a correct decision regardless of which algorithm it adopts. Furthermore, when the number of

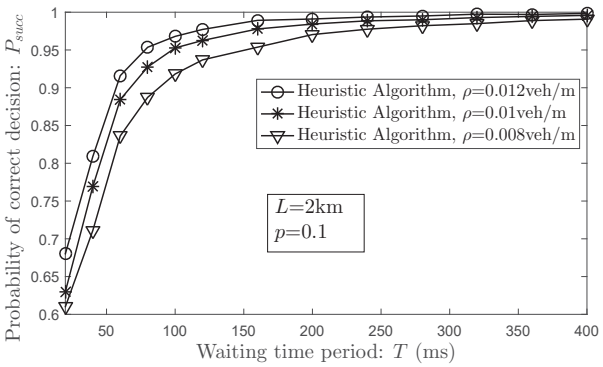

Fig. 3. An illustration of the relationship between the probability of correct decision and the waiting time period the destination vehicle waits before it starts to make a final decision by adopting the proposed algorithm.

malicious vehicles in the network reaches a certain threshold, most of the message dissemination paths will be compromised. In that case, the destination vehicle will be totally misguided by the incorrect messages and the message security performance approaches its minimum value $P_{succ} = 0$.

Fig. 3 demonstrates the relationship between the probability of correct decision P_{succ} achieved assuming our proposed algorithm, and the waiting time period T the destination vehicle waits before it starts to make a decision. Importantly, we can see that a longer waiting time is beneficial to the secure message dissemination because it potentially implies a larger number of received messages. This consequently, brings more information on the underlying network topology, and therefore leads to a more robust result of the data consistency check. However, when T increases beyond a certain threshold T_{th}, e.g., in the case of ρ=0.01veh/m, $T_{th} = 150ms$, a further increase in T has only marginal (less than 5%) impact on the probability of correct decision. This is due to the fact that when T is larger than a threshold, the marginal return brought by waiting a longer time to the security performance is diminishing. Furthermore, it can be seen that to achieve the same message security performance, when the vehicular density is lower, the waiting time needs to be longer. Therefore, when determining the waiting time period, it is important to take the vehicular density into account, e.g., in areas where the vehicular density is large, the waiting time can be reduced. Thus, Fig. 3 helps to provide guidance on the choice of waiting time period for destination vehicles.

V. Conclusions

This paper proposed a novel heuristic decision algorithm that utilizes the underlying network topology information to address the issue of message inconsistency caused by malicious vehicles that would tamper the content of disseminated messages. The heuristic decision algorithm proposed enables a vehicle to make a decision when receiving conflicting messages purely based on network topology information, without the need for a prior knowledge of the percentage of malicious vehicles in the network, and therefore, is fairly easy to implement in practice. By comparing the proposed algorithm with existing algorithms that do not consider the underlying topological information or only partially consider message correlation, we showed that our proposed algorithm greatly outperforms existing ones. Moreover, we discussed the impact of some key parameters on the performance of the algorithm, including the percentage of malicious vehicles in the network, and the waiting time the destination vehicle waits before making the decision. Our results give insight on the optimum decision algorithm design for vehicular networks to improve message security.

References

[1] H. Tan, M. Ma, et al., "A secure and authenticated key management protocol (SA-KMP) for vehicular networks," *IEEE Trans. Veh. Technol.,* vol. 65, no. 12, pp. 9570–9584, Dec. 2016.

[2] J. Petit, F. Schaub, et al., "Pseudonym Schemes in Vehicular Networks: A Survey", *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 228-255, First Quarter, 2015.

[3] Q. Yang, H. Wang, "Toward trustworthy vehicular social networks", *IEEE Commun. Mag*., vol. 53, no. 8, pp, 42 - 47, Aug. 2015.

[4] S. Ahmed, S. Al Rubeaai, et al., "Novel Trust Framework for Vehicular Networks", *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9498 - 9511, Oct. 2017.

[5] Y. Du, M. Chowdhury, et al., "A Distributed Message Delivery Infrastructure for Connected Vehicle Technology Applications", to appear in *IEEE Trans. Intell. Transp. Syst.*, 2017.

[6] S. Dietzel, J. Petit, et al., "Graph-Based Metrics for Insider Attack Detection in VANET Multihop Data Dissemination Protocols," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1505-1518, May. 2013.

[7] M. Raya, P. Papadimitratos, et al., "On Data-Centric Trust Establishment in Ephemeral Ad Hoc Networks," in *IEEE INFOCOM*, 2008.

[8] J. Radak, B. Ducourthial, et al., "Detecting Road Events Using Distributed Data Fusion: Experimental Evaluation for the Icy Roads Case," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 184-194, Jan. 2016.

[9] M. A. Javed, E. B. Hamida, "On the Interrelation of Security, QoS, and Safety in Cooperative ITS", *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp 1943 - 1957, Jul. 2017.

[10] J. Ponniah, Y. C. Hu, et al., "A Clean Slate Approach to Secure Ad Hoc Wireless Networking - Open Unsynchronized Networks", *IEEE Trans. Control Netw. Syst.*, vol. 4, no. 1, pp. 37 - 48, March, 2017.

[11] M. H. DeGroot, and M. J. Schervish, Probability and statistics, fourth edition, Boston, MA: Addison-Wesley, 2002.

[12] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2. New York, NY, USA: *Wiley*, 1971.

[13] J. L. Gross, and J. Yellen, Graph theory and its applications, second edition, Boca Raton, Florida, USA, Chapman & Hall/CRC, 2006.

[14] D. R. Karger and C. Stein. A new approach to the minimum cut problem. J. ACM, vol. 43, no. 4, pp. 601–640, Dec. 1996.

[15] Y. Wang, J. Zheng, and N. Mitton, "Delivery Delay Analysis for Roadside Unit Deployment in Vehicular Ad Hoc Networks with Intermittent Connectivity", *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8591 - 8602, Oct. 2016.

[16] Z. Zhang, G. Mao, and B. D. O. Anderson, "Stochastic Characterization of Information Propagation Process in Vehicular Ad hoc Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 122-135, Feb. 2014.

[17] Y. Zhu, Multisensor Decision and Estimation Fusion, United States, *Springer*, 2003.

[18] Z. Huang, S. Ruj, et al., "A social network approach to trust management in VANETs," *Peer-to-Peer Networking and Applications*, vol. 7, no. 3, pp. 229-242, 2014.