Host Based Intrusion Detection for VANETs: A Statistical Approach to Rogue Node Detection

Kamran Zaidi, Milos Milojevic, *Student Member, IEEE,* Veselin Rakocevic, *Member, IEEE,* Arumugam Nallanathan, *Senior Member, IEEE,* Muttukrishnan Rajarajan, *Senior Member, IEEE*

***Abstract*—In this work, an Intrusion Detection System (IDS) for vehicular ad hoc networks (VANETs) is proposed and evaluated. The IDS is evaluated by simulation in presence of rogue nodes that can launch different attacks. The proposed IDS is capable of detecting a false information attack using statistical techniques effectively and can also detect other types of attacks. First, the theory and implementation of the VANET model that is used to train the IDS is discussed. Then an extensive simulation and analysis of our model under different traffic conditions is conducted to identify the effects of these parameters in VANETs. In addition, the extensive data gathered in the simulations is presented using graphical and statistical techniques. Moreover, rogue nodes are introduced in the network and an algorithm is presented to detect these rogue nodes. Finally, we evaluate our system and observe that the proposed application layer IDS based on cooperative information exchange mechanism is better for dynamic and fast moving networks such as VANETs as compared to other techniques available.**

***Index Terms* -** Intrusion Detection, Security, wireless networks, cryptography, rogue nodes, fault tolerance, VANETs, vehicular networks.

I. INTRODUCTION

VANETs are considered to be the next big thing that will change our lives remarkably. It is only logical that technology is used to make our lives and roads safer. The automotive industry looks all set to equip their vehicles with Wireless Access Vehicular Environment (WAVE) devices from 2015, this will enable vehicles to communicate with each other to exchange safety information. Moreover, autonomous vehicles are not that far off either with Google Car a reality today. These technological innovations in our vehicles will change the way we think about road travel by making it much safer and productive. WAVE protocols are based on IEEE 802.11p standard and provide the basic radio standard for Dedicated Short Range Communication (DSRC) in VANETs. Vehicles use DSRC to communicate with each other i.e. vehicle to vehicle (V2V) and with the infrastructure (Road Side Units - RSUs) i.e. vehicle to infrastructure (V2I) communication.

VANETs will become a reality in the very near future. The tremendous safety, convenience and commercial potential of vehicular networks will not only drive its deployment but will be fuelled by its demand as well once consumers realize its effectiveness. VANETs have the ability to make roads safer especially in conditions which are currently considered hazardous and unavoidable. Imagine the ability to be able to navigate safely in otherwise very dangerous driving conditions like fog, accidents, black ice. However, there are some very serious security issues that need to be addressed before the full potential of VANETs can be realized. Vehicular networks are very fast moving and highly dynamic due to which it is very important that the information being shared is authentic and actionable. As encounters will be short lived and the received information has to be actioned quickly, therefore, it is important that the reliability of the information is ascertained quickly.

In ad-hoc networks, maintaining and depending on trust or reputation is very expensive and a complex concept. In VANETs, centralized trust has long been debated as it is difficult to maintain, update and use. The existing mechanism for authenticating messages in vehicular networks involves the use of cryptography [7]–[9] and trust [18]–[20]. Cryptographic techniques involve paired keys and overhead in terms of computing cost, storage and time. Even with cryptographic techniques, security lapses are inevitable leading to intrusions due to stolen keys or compromised Trusted Authorities etc. An attack is especially difficult to prevent when it is launched from within the network. Due to the wireless and mobile nature of vehicular networks and its dynamic topology, it is not possible to use the same intrusion detection mechanisms that are used in wired networks. Therefore, it is essential that an intrusion detection system is deployed to detect attacks and help secure VANETs. The proposed IDS will detect different types of attacks launched by rogue or compromised nodes in the network. The IDS will then be able to minimize the damage to the network by taking necessary actions. The proposed IDS works in a distributed manner and is designed for deployment at each host node in the vehicular network.

A. Our Contributions

The main contributions in this paper are:

1) An Intrusion Detection System is proposed that uses statistical techniques to detect anomalies and identify rogue nodes using a traffic model. We extend the earlier work done in [23] significantly by extensive simulations under varying vehicular and network traffic conditions

Kamran Zaidi, Milos Milojevic, Veselin Rakocevic and Muttukrishnan Rajarajan are all with School of Mathematics, Computer Science and Engineering, City University London, London, EC1V 0HB, UK. (Emails: (kamran.zaidi.1, milos.milojevic.1, veselin.rakocevic.1, r.muttukrishnan)@city.ac.uk)

Arumugam Nallanathan is with Department of Informatics at King's College London, London, WC2R 2LS, UK. (Email: arumugam.nallanathan@kcl.ac.uk)

and using statistical techniques to determine false data especially in emergency messages.
2) The extensive data collected is analysed using statistical techniques and the decision to accept or reject data is based on hypothesis testing.
3) The effects of various parameters such as transmission intervals and vehicle density are also shown.
4) The proposed IDS is not dependent on any infrastructure such as RSU or expensive hardware such as Lidar, radar or cameras.
5) Using the proposed mechanism the network message congestion is controlled by reduced message transmissions i.e. prevents broadcast storms. Moreover, we show that using the proposed model and IDS it is possible for vehicles to keep the network functioning even when up to 40% nodes are malicious and contribute false parameter values.

The rest of this paper is organized as follows: related work is discussed in Section II. The system and the attack model is presented in Section III. In Section IV, overview of the proposed IDS is presented. Section V evaluates the security performance of the proposed IDS in detail. The results are discussed in Section VI and the conclusions and future work are given in Section VII.

II. RELATED WORK

Security of VANETs is a very important issue and has been the focus of research for the last many years. The vehicular networks are unique as the users will be making life saving decisions based on the information being received. It is therefore, imperative that there is a mechanism to detect false information. Researchers have proposed using Cryptography and digital signatures to secure and sign messages so as to ensure integrity and non-repudiation of messages in VANETs. Digital signatures have been proposed for VANETs in [4]–[6]. Different schemes have been proposed including Public Key Infrastructure (PKI) [7]–[9].

The propagation of emergency messages in VANETs is done either through multi-hop or by broadcasting them. Therefore, malicious behaviour e.g. false information attack is possible even in case of strong cryptography as insiders can turn malicious. A malicious user might send false alert to clear the road for himself or cause havoc by creating a traffic jam by sending a fake accident alert. Researchers in [13] suggest using data centric techniques to make VANETs more reliable by only considering the data being shared. For fast moving and dynamic networks information centric schemes are required in addition to the cryptography and certificates to protect against inside attacks.

There are mainly two approaches for dealing with the false information attacks i.e. Trust or Reputation based schemes and Data Centric Schemes. This trust based on reputation can either be infrastructure based or self organising [17]. Self organising trust means to assign a trust score to another user based on previous or current interactions. This trust score represents the reputation of the user in the network and helps other nodes decide whether it can be trusted or not. Such voting schemes (credit scores) are promising in wired networks or online systems where the users have a fixed physical identity but difficult to implement in a fast moving and rapidly changing network such as VANETs. Reputation based schemes have been proposed in [18]–[20]. In [19], [20] a decentralized infrastructure has been adopted whereas in [18] a centralized infrastructure is proposed. Reputation and Trust based schemes are useful but cannot be used for detecting false emergency messages as trust is built over a period of time and if a false message comes from a trusted node then there is no way to detect it.

Data centric misbehaviour detection techniques have been proposed in [15], [18]. In [18] the authors propose a model of VANETs to be used to detect and correct errors in the data being sent out by vehicles. The messages that conform to the model are accepted and rejected otherwise. In [15] emergency messages are relayed and false information is identified based on the kind of message and the subsequent behaviour of the sending vehicle. Such a technique will not be feasible for emergency messages which need to be acted on quickly. Also, such a scheme will increase the computation cost for the nodes. A misbehaviour detection system and eviction mechanism is proposed in [16] where nodes are termed misbehaving if their info is inconsistent with the situation. Once a node is classified as misbehaving node then the neighbouring nodes can temporarily evict them by sharing warning messages about them and later their credentials are passed on to the CA which revokes them by adding them to a Revocation List (RL). However, as discussed previously the RLs are difficult to manage and use in VANETs.

Intrusion detection is the most reliable approach to protect vehicular networks against threats as it has the ability to detect insider and external attacks with a high accuracy [2]. Some research has been done in the area of IDS / IPS for Mobile Ad-hoc Networks (MANETs) and VANETs in [1], [24]–[30]. In [26], the authors propose an acknowledgement scheme to prevent packet dropping and false misbehaviour report generation by nodes for MANETs to report or convict a rogue node. In [27], the authors propose a watchdog for intrusion detection in VANETs. The watchdog works by monitoring all packets to decide if an attack is under progress. In [25], trust and position information is combined to determine if a vehicle is falsifying its position i.e. if the position claimed by one vehicle overlaps the position claimed by another in which case the vehicle with low trust value is flagged as an intruder. In [24], a method is proposed to detect intrusions through trust by assigning reputation scores to vehicles and the RSUs are used to compute these scores and the CA aggregates them. Similarly, in [1], rule based anomaly detection and reputation scores are used for the IDS in vehicular network. In [28], [29], intrusion prediction approaches have been discussed.

Intrusion detection systems are very effective as they are able to detect attacks from insiders at real time but at the same time need to be updated for new attacks. Moreover, IDS need strong authentication and identification systems in order to work properly. Intrusion prediction systems on the other hand try to predict new attacks that can protect the system

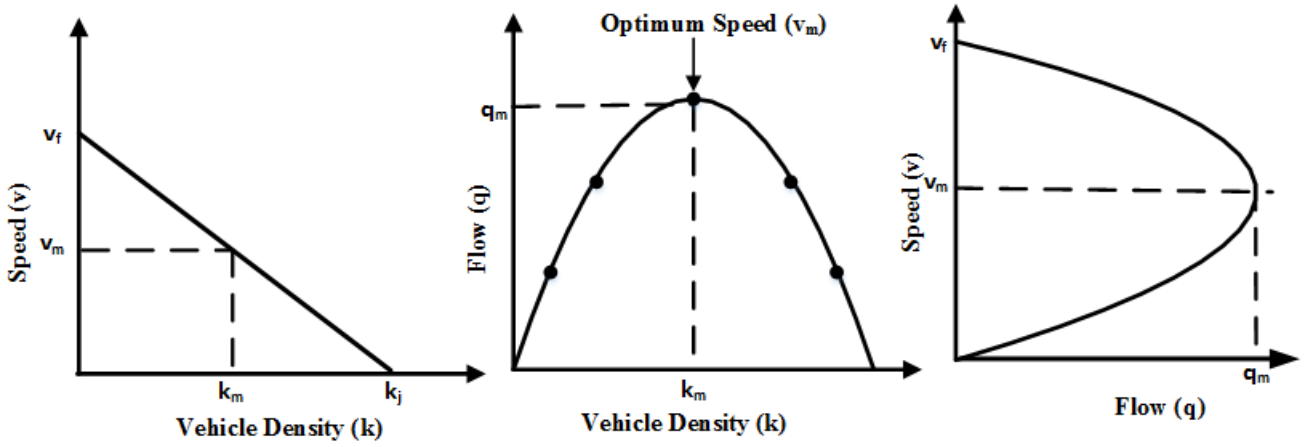


Fig. 1: Greenshield's Fundamental Diagrams (a)Speed vs Vehicle Density, (b)Flow vs Vehicle Density, (c)Speed vs Flow

from unknown attacks. However, the probability thresholds need to be set carefully in such intrusion prediction systems to get accurate results. This work proposes an IDS that does not use trust or reputation and only relies on the analysis of the received data to detect intrusions in the network. The statistical technique used in the IDS declares data true or false which leads to the node being declared honest or rogue instead of the other way around.

III. PRELIMINARIES

A. Authentication & Privacy Preservation

In any network, it is very important that nodes can be identified correctly and are distinguishable from one another but at the same time privacy is preserved. This means that all nodes are authenticated by a Certificate Authority (CA). It is assumed that all vehicles have authenticated themselves with a certificate authority and obtained a valid certificate and public/private key pairs (Pseudonyms-PNs). The keys are used to encrypt the routine messages and others can authenticate and decrypt the messages by using the relevant public keys. It is also assumed that all vehicles have enough key pairs to last them a long time and they keep changing these keys to preserve their privacy. However, these keys are in a reasonable time i.e. not too quickly to avoid short term linkability. This ensures that even by changing PNs the recent messages of a node can be linked to the same node. Therefore, the proposed IDS allows the nodes to change their PNs but can still keep track of the rogue nodes.

B. VANET Model

In order to model the flow of traffic on motorways / highways a mathematical model is needed. Therefore, the Greenshield's model which is considered to be a fairly accurate model in traffic engineering for estimating and modelling uninterrupted traffic (without traffic signals) is utilized. Greenshield's model uses standard parameters such as flow (vehicles per hour) and density (vehicles per km). The model describes the relationship between speed (v) and density (k) of vehicles as being negatively correlated with density increasing with the decrease in speed as shown in Fig 1(c). In the figure v_f is the free flow speed when density is zero i.e. vehicles can choose to move freely as there are no or very few vehicles on the road. As the density of vehicles increases the speed decreases till density reaches the maximum which is referred to as jam density or k_j at which point the speed becomes zero and vehicles are stuck in a traffic jam. In the figure k_m and v_m are the optimal density and speed respectively which allows the traffic to progress at the optimum rate of flow - q_m Fig 1 (a), (b) & (c). The flow is given as:

$$q = k \times v \tag{1}$$

The relationship between speed and density is given as:

$$v = v_f - \frac{k}{k_j} v_f \tag{2}$$

From (1) & (2) the relationship between speed and density can be found to be:

$$q = v_f k - \frac{k^2}{k_j} v_f \tag{3}$$

Each vehicle can calculate the density of vehicles on the highway around it by the number of messages it receives from other vehicles by checking their IDs from messages. This enables each vehicle to calculate the density quite accurately in a moving window around itself as shown in Fig 2. The size of this density window is equal to the transmission and reception range of a vehicle (500 meters). This means that a vehicle can receive messages from a vehicle which is up to 500m ahead of it and 500m behind it. Therefore, each vehicle has a communication window of 1000m around it that it can use to calculate the density ($Density_{calc}$). Also, each vehicle can calculate the average speed of vehicles ($Speed_{AVG}$) within its communication window. In our scheme each vehicle transmits not only its location and speed but the calculated value of flow as well. Therefore, the vehicles calculate the traffic flow parameter using density and average speed of other vehicles through Greenshields model. The flow serves as a global parameter which each vehicle calculates on its own and should be very similar for vehicles that are close to each other in the same traffic conditions. Moreover, information will be

considered correct if it conforms to this model and false otherwise.

The idea behind this mechanism is that in case of an emergency (an accident or sudden braking) all vehicles behind the incident will apply brakes and therefore, their flow values will go down. These low values of flow will be transmitted to other vehicles behind them which will cause their calculated flow values to go down as well as shown in Fig. 2. The red region is where the brakes have been applied, the orange region is where the effect information is being propagated and they are getting information of an accident up ahead. The blue region is some distance away where vehicles are getting reports of some congestion ahead on the highway but they don't have to start braking just yet. This is one of the benefits and desirable effects of the proposed model, as there is no need to flood the network with the congestion warning and instead the information is propagated gracefully. However, in case of a false emergency message; a vehicle will try to create the illusion of an accident by lowering its flow and speed values and transmitting it to others. However, as there is only one vehicle that is transmitting this low value, it can easily be flagged and identified.

Each vehicle transmits its $Flow_{AVG}$ which becomes $Flow_{Rcvd}$ for other vehicles. If a vehicle receives a value of Flow from another vehicle that does not agree with the VANET model then the data is rejected and vehicles' ID is noted and reported. If the data agrees with the model then the receiving node checks the data with its own calculated values to confirm its values are indeed correct. If the values do not agree with the node's own calculated parameters of Flow, Speed and Density then the values are discarded and the sender ID is reported. The two values of flow are calculated as follows:

$$Flow_{OWN} = Speed_{AVG} \times Density_{calc} \tag{4}$$

$$Flow_{AVG} = \frac{1}{n}\sum^{n} Flow_{Rcvd} \tag{5}$$

C. Message Format

Each vehicle creates its own message m for beacon and apart from the usual parameters also includes the following:

$$m(Speed_{own}, Density_{calc}, Flow_{AVG})$$

Each beacon message m is hashed ($H(m)$) and signed by the vehicle using its secret key (SK).

$$sig = SK(H(m))$$

The details of how this signature is generated and how they are verified are not in the scope of this paper. In case of emergency e.g. an accident or emergency braking, each vehicle generates an emergency message which has the following format:

$$EmergencyMsg(Type, Flow_{AVG}, Speed_{Own}, Density_{calc})$$

where field - $Type$ can be Emergency Braking, Accident Ahead, Slippery Road etc. It must be noted that the emergency messages are not encrypted and have to be actioned quickly by those receiving it.

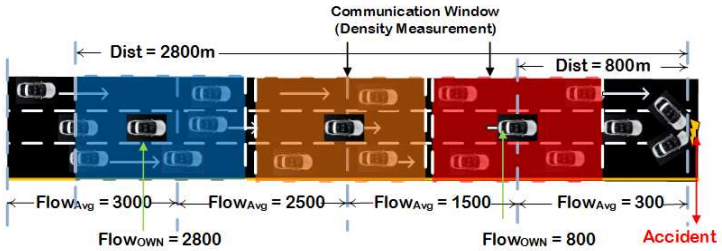


Fig. 2: Decreasing value of Flow in case of accident

D. Attack Model

There are different types of attacks that can take place in VANETs. We will be looking at the following attacks:

1) *False Information Attack:* A rogue node can inject false data in the network either on purpose with malicious intent or due to faulty sensors that can cause serious damage to the network. Under extreme conditions the network can even be paralysed. The rogue node can start injecting false data at any time and can falsify values of their own speed and their calculated values of flow and density either in beacon message or emergency message. In case of a false emergency message the rogue node will start sending a low value of Flow or sudden decrease in speed or both to indicate an accident or emergency braking.
2) *Sybil Attack:* Another attack that a rogue node can launch is a Sybil attack i.e. when a rogue vehicle transmits multiple messages each with a different ID to indicate that it is not one vehicle but many vehicles thereby giving a false impression of congestion by lowering the Flow values in the messages. The IDs could either have been spoofed or stolen from compromised nodes.

IV. IDS OVERVIEW

The host based Intrusion Detection System proposed in this work is deployed at each vehicle and is able to detect intrusions in VANETs and then take corrective measures to contain the damage. In order to train the IDS, a model of the network under normal conditions is needed so that deviations (anomalies) from the normal behaviour can be detected and alarms can be raised i.e. other vehicles can be informed (shown in Fig. 3). In the proposed model discussed in the previous section, the vehicles send their speed, calculated average flow, calculated density and location information to other vehicles. Also, each vehicle calculates its own value of average flow which provides them with a very good estimate of the traffic in their vicinity and up ahead as well.

A. Cooperative Data Collection

Using our scheme each node (vehicle) collects data from other nodes (vehicles) in its vicinity to model the traffic around it. The vehicles cooperate with each other and share the values of their parameters using the Greenshield's model described above. As a vehicle will receive the parameter values from

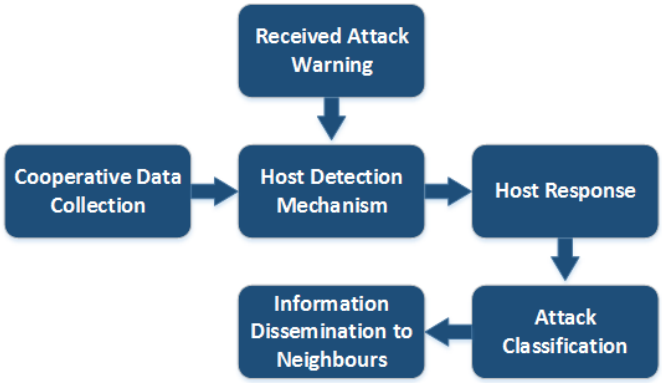


Fig. 3: Proposed Host Based Intrusion Detection System

all other vehicles within range, therefore, each vehicle has information about all the vehicles in that region. Due to this each vehicle can calculate the (estimate) mean μ. The trace data has shown that under all conditions the flow values will be close together and will lie within two standard deviations of the mean. This means that all vehicles that are within communication range are calculating very similar value of the $Flow_{AVG}$ as they are under similar traffic conditions. This is obvious as all nodes are dependent on other nodes to calculate their parameter values in all circumstances i.e. free flowing traffic and in case of an accident. When enough readings / data has been gathered, the conditions of the central limit theorem apply and we approach a normal distribution. To show this we plot the frequency distribution of the Average Flow Values $Flow_{AVG}$ of a random node e.g. Node No. 90 in our simulation with vehicle inter-arrival time of 2 sec, transmission interval of 0.5 sec from simulation time t=203 sec to t=325 sec as shown in Fig 4. The data is slightly left skewed as vehicles start from rest and therefore, have lower values of flow in the beginning. This means that we are now in a position to set up a hypothesis test and use the t-test for detecting false values reported by a rogue / malicious vehicle. The t-test for comparing the two population means is used as the sample size can be small.

The parameter values follow a normal distribution and as the received values are in pairs, therefore, we use the paired t-test to calculate the probabilities associated with getting values in different ranges. The standard deviation and the test statistic t_o are calculated as:

$$t_o = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_1} + \frac{s_y^2}{n_2}}} \tag{6}$$

Here, $\bar{x}$ is the mean difference of the received values and $\bar{y}$ is the mean difference of vehicle's own calculated values, s_x and s_y are the standard deviations of received and vehicles own calculated values respectively. n_1 and n_2 are the number of samples for the received and own values respectively. The degrees of freedom will be $n_1 + n_2 - 2$. The algorithm of the proposed IDS is given in Algorithm. 1, the data is collected from all neighbouring nodes and checked if there is a significant difference between the calculated and received values. If there is a significant difference then the node is monitored and some parameter values are collected (accepted) initially. Once sufficient samples have been collected then the t-test is carried out. If the t-test gives a result within the acceptance region then the data is accepted and else rejected. If the data is rejected then the node is highlighted as rogue and the attack is classified as Information Attack and subsequent values from that node are rejected. A message is then sent to other users informing them of the rogue node and the type of attack being launched by that node.

Algorithm 1 Algorithm for IDS

```
for each msg received do
    Update Density_calc
    Update Speed_AVG
    Flow_OWN = Speed_AVG x Desnsity_calc
    if Flow_Rcvd − Flow_OWN < Threshold then
        Accept Data
        Calculate Flow_AVG
    else
        Monitor Node and Accept Data temporarily
        if Hypothesis Test == Reject then
            Reject Data
            Report Node
            Calculate Flow_AVG
        end if
    end if
end for
```

B. Hypothesis Testing for Data Correctness

Hypothesis testing is a common technique used in engineering applications to test two claims when only one of them can be true. The hypothesis testing approach also assigns a confidence interval to a range of values that enables us to accept a claim with a certain confidence. This suits us as in our VANET model and proposed IDS there are two possibilities i.e. either the node is honest and we accept its data or the node is rogue and we reject its data. To check whether hypothesis testing works well in our model, we ran the simulations numerous times in OMNET++ and then exported the data to MS Excel and Matlab to analyse and visualize it.

We use hypothesis testing to decide whether a received parameter value should be accepted or not. If the received value is within the 99% confidence interval i.e. within the

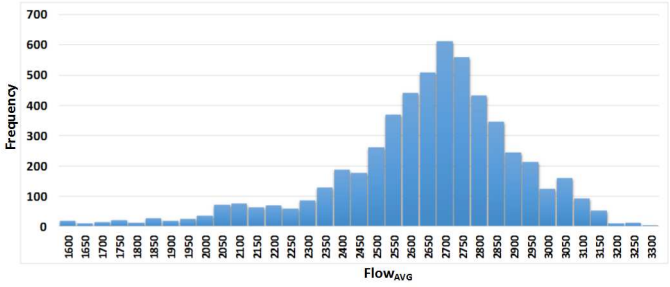


Fig. 4: $Flow_{AVG}$ values for Node 90 between t=203 sec to t=325 sec

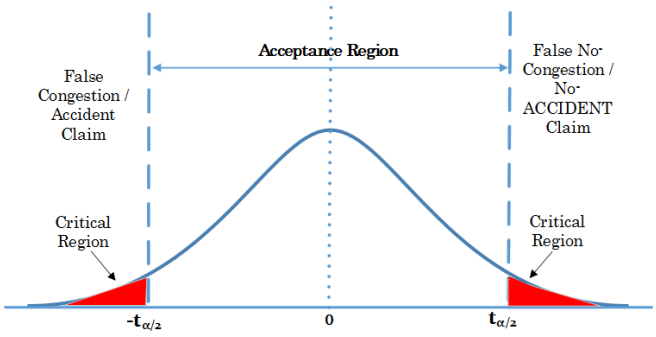


Fig. 5: Distribution of t_o for $Flow_{AVG}$

acceptance region, then the value is accepted. If the received flow value is within the rejection region then it is rejected. This is shown in Fig 5. There are always two hypotheses stated, there is the null hypothesis H_o which we want to test (and assumed to be correct) and alternate hypothesis H_a. If the null hypothesis is rejected then the alternate hypothesis is accepted and if we do not have enough evidence against the null hypothesis then it is accepted. The null hypothesis H_o in our case is that the data (Flow value) received is from an honest node. The alternate hypothesis H_a is that the value received is false (from a rogue node) and we fail to accept (reject) it. In other words we say that we don't have enough evidence to accept the received data and therefore, we reject it. The Hypotheses that will be tested in the host IDSs are stated below:

H_o : Accept Received data (Node is Honest)
H_a : Reject (Fail to Accept) Received data (Data is false & Node is Malicious or Rogue)

The IDS in each vehicle also computes a p-value that helps it in accepting or rejecting the null hypothesis. The p-value gives the probability of getting a value which is atleast as extreme so, the p-value gives information about the weight of evidence against the null hypothesis H_o i.e. the smaller the p-value the greater the evidence against H_o. There are two types of errors associated with hypothesis testing as shown in Table I. In our scenario, Type-2 error (false negative) is not very serious as the worst case scenario is slowing down whereas Type-1 error (false positive) is very serious. Therefore, keeping this in view we use a wide confidence interval. The level of significance is denoted by α. The usual values of α are taken to be 0.01(1%) or 0.05(5%) which means the probability that the test statistic falls in our acceptance region is $1-\alpha$ and the confidence interval for the two values of α = 0.01 and 0.05 are 99% and 95% respectively. We take the value of α to be 0.01 and as this will be a two-tailed test therefore, the upper and lower limit of our acceptance region will be $t_{\alpha/2}$ & $-t_{\alpha/2}$ as shown in Fig. 5. The degrees of freedom will be $n_1 + n_2 - 2$

TABLE I: Decisions in Hypothesis Testing

	Node is Honest - H_o	Node is Rogue - H_a
Accept H_o	No Error	Type 2 Error
Reject H_o	Type 1 Error	No Error

TABLE II: SIMULATION PARAMETERS

PARAMETER	VALUE
Simulation Time	400 sec
Scenario	3 Lane Highway
Highway Length	5-Kms
Max Vehicle Speed	28 m/sec or 100 Km/hr
Mobility Tool	VACaMobil
Network Simulation Package	OMNET++
Vehicular Traffic Generation Tool	SUMO
Vehicle Inter-Arrival Rate	1s, 2s and 3s
Transmission Rate	Every 0.2s, 0.5s and 1s
Wireless Protocol	802.11p
Transmission Range	500m in each direction

and the corresponding limits can be looked up from the t-table. This means that the probability is α that the test statistic t_o falls in the region $t_o > t_{\alpha/2}$ or $t_o < -t_{\alpha/2}$ when the null hypothesis H_o is true. Therefore, we will reject the received value if it is outside the acceptance region i.e. we reject the value if either:

$$-t_{\alpha/2} > t_o > t_{\alpha/2}$$

In our case the received flow values for any chosen node are always within the acceptance region or within the 99% confidence interval as long as the node is honest. In the case of an accident as the values will drop, they will have an impact on all vehicles in the region which will bring down the $Flow_{AVG}$ value for the region and as a result the values are still within the acceptance region as the standard deviation increases.

As shown in Fig. 5, there are two cases where the rogue node will falsify its values i.e. it can either deny congestion or accident or it can wrongly give the impression of congestion or accident. Therefore, the IDS can decide which category the false information falls under depending on whether $t_o > t_{\alpha/2}$ or $t_o < -t_{\alpha/2}$.

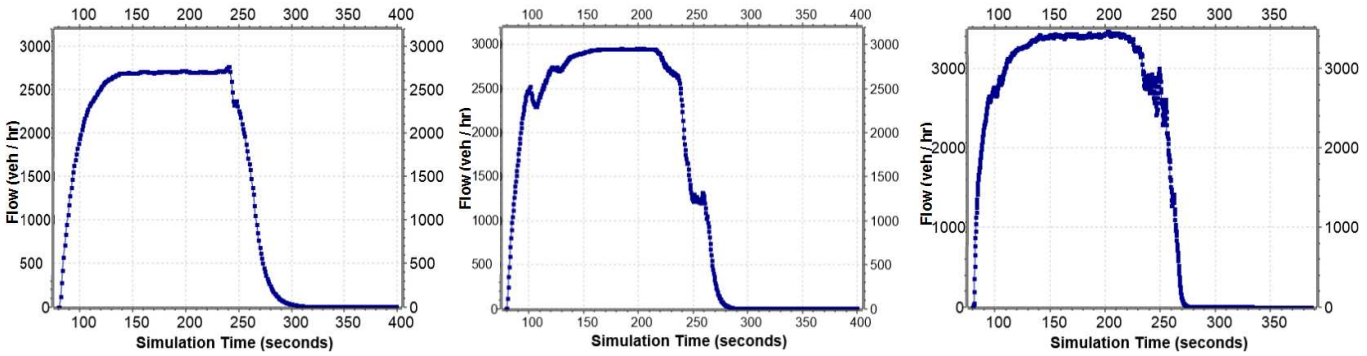
V. PERFORMANCE EVALUATION

A. Simulation Setup

In order to check the proposed IDS extensive simulations were done using OMNET++, SUMO [22] and VACaMobil [21]. OMNET is a modular C++ library and framework that is used for network simulations. Simulation of Urban Mobility (SUMO) is a software tool used to generate vehicular traffic by specifying speed, types, behaviour and number of vehicles. Sumo also sets up road types and conditions. VACaMobil is a car mobility manager for OMNET that works in parallel with SUMO.

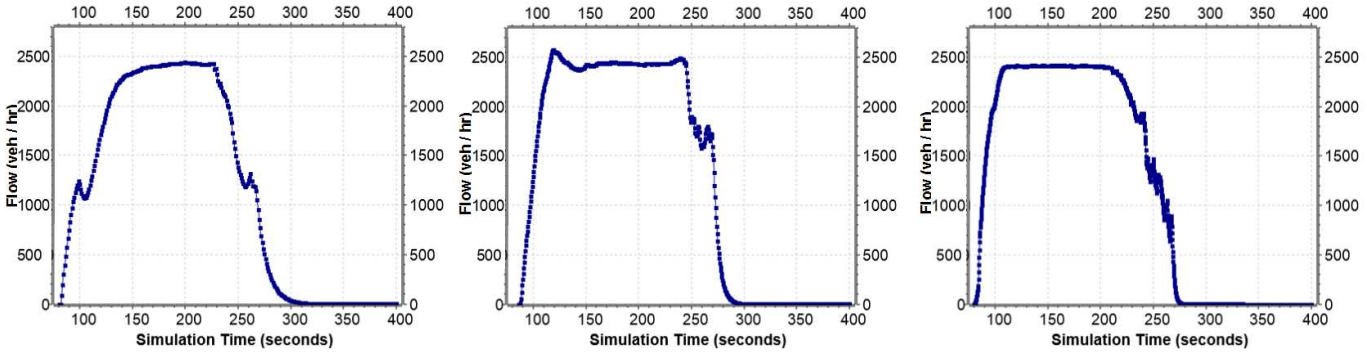
The scenario is simulated with parameters shown in Table II. In order to gather data for anomaly detection we use different scenarios. We gather data when there is no accident and no rogue nodes to understand and develop the model under normal circumstances. Data is also collected for runs in case of an actual accident to understand how parameters will change. Furthermore, rogue nodes are inserted in both cases i.e. in case of normal conditions (no-accident) and in case of an actual accident to see how well our IDS works. The simulations are carried out with varying values of the following parameters:

1) ***Density:*** The density of nodes is an important parameter for ad-hoc networks and especially for VANETs. As



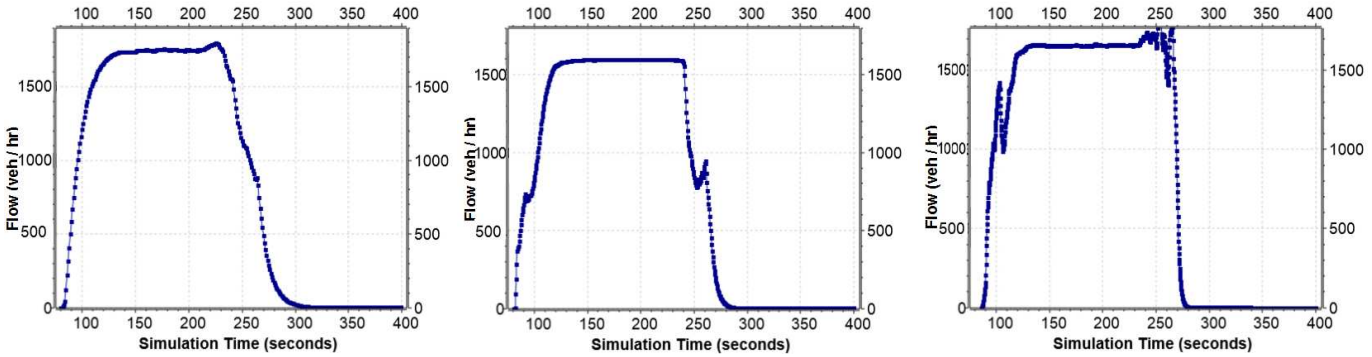
(a) Node 50, Update Interval 1sec (b) Node 59, Update Interval 0.5sec (c) Node 56, Update Interval 0.2sec

Fig. 6: Accident Scenario: Inter-Arrival time = 1 sec: All Vehicles starting at approx t = 80sec



(a) Node 39, Update Interval 1sec (b) Node 40, Update Interval 0.5sec (c) Node 36, Update Interval 0.2sec

Fig. 7: Accident Scenario: Inter-Arrival time = 2 sec: All vehicles starting at approx t = 80sec



(a) Node 32, Update Interval 1sec (b) Node 26, Update Interval 0.5sec (c) Node 24, Update Interval 0.2sec

Fig. 8: Accident Scenario: Inter-Arrival time = 3 sec: All vehicles starting at approx t = 80sec

the channel bandwidth is limited, it is essential to keep it under consideration and observe its effects on any system. In this work, we vary the density of vehicles by changing their inter-arrival time i.e. the time that they are inserted in the simulation. We use OMNET's exponential inter-arrival distribution with a time of 1, 2 and 3 seconds.

2) ***Beaconing Rate or Sampling Rate:*** This is the beaconing time period after which each vehicle is transmitting its parameters to other vehicles. We have used variable time periods to observe the effects of this on VANETs in general and the proposed IDS in particular. We have used time periods of 0.2, 0.5 and 1 seconds. It is worth mentioning that the recommended beaconing rate in IEEE 809.11p is a 100 milliseconds (0.1 sec). The minimum time period of 0.2 seconds was chosen as the generated data set was becoming too large and data processing was becoming a problem.
3) ***Number of Rogue Nodes:*** The number of rogue nodes is varied to evaluate the performance of the proposed scheme and the IDS in these circumstances.

A large amount of trace data is generated with the simulation runs by varying parameters described above. For example, the most data generated in this work in one simulation is if the

sampling rate is 0.2 sec, the total number of vehicles which are active in simulation in case of an accident, are 300 and the simulation time is 400 seconds then more than 18000 data points are generated and collected out of which around 10,000 are vectors. The minimum data generated in one simulation is when the sampling rate is 1 sec, the total number of vehicles which are active in simulation in case of no accident are around 150 and the simulation time is 400 seconds then around 10,000 data points are generated out of which around 6,000 are vectors. The parameters of interest from the large data set were exported to MS Excel and Matlab for analysis, testing and visualization.

B. Simulation Results

1) Actual Accident Scenario - No Rogue Nodes: The results for the actual accident scenario are shown in Figs. 6, 7 & 8. The density of vehicles (controlled by Inter-Arrival Time) and the update interval (transmission rate) are varied in the simulations to study their effects. What is noteworthy here is that the flow parameter gradually decreases which proves our earlier assumption.

In Fig. 6 (a), (b) & (c) the results are shown for the value of $Flow_{AVG}$ for vehicles that are starting at approximately t=80 sec and an accident occurs at t=180 secs for the same density of vehicles. Similarly, Figs. 7 & 8 show the results when the density is kept constant and the update interval is varied. It can be seen from Figs. 6, 7 & 8 that the density has a negligible effect on the working of the method i.e. all vehicles receive the information about the attack at the same time (i.e. Figs in the same column such as 6b, 7b, 8b) if the update interval is the same. This shows that the proposed mechanism is scalable. Also, it is clear that the update interval has a significant impact on the information flow as the value settles down the quickest in Figs 6c, 7c, 8c) when the update interval is the smallest i.e. 0.2 sec as compared to the others when the update interval is higher. However, this is acceptable as the standard update interval in VANETs can be as low as 100 msec or 0.1 sec.

2) Normal Traffic - No Accident - No Rogue Nodes: In order to record the traffic data in case of normal traffic i.e. no accident and no rogue nodes to see how the system works. Fig. 9 shows the recorded data for the 100th node when update interval is 1sec and inter-arrival rate is 1 sec. As expected, the average value of Flow ($Flow_{AVG}$), calculated values for flow ($Flow_{OWN}$) & the received flow values from other vehicles ($Flow_{RCVD}$) are all quite close to each other and the received values ($Flow_{RCVD}$) are in fact within one standard deviation of the ($Flow_{AVG}$) as calculated by the node.

3) No Accident - Rogue Nodes: A scenario is simulated in which there is no accident but rogue nodes start transmitting a low false value of Flow after t=160 sec. We run the simulations both with and without the proposed IDS and also vary the number of rogue / malicious nodes and collect the data. The results are shown with and without the proposed

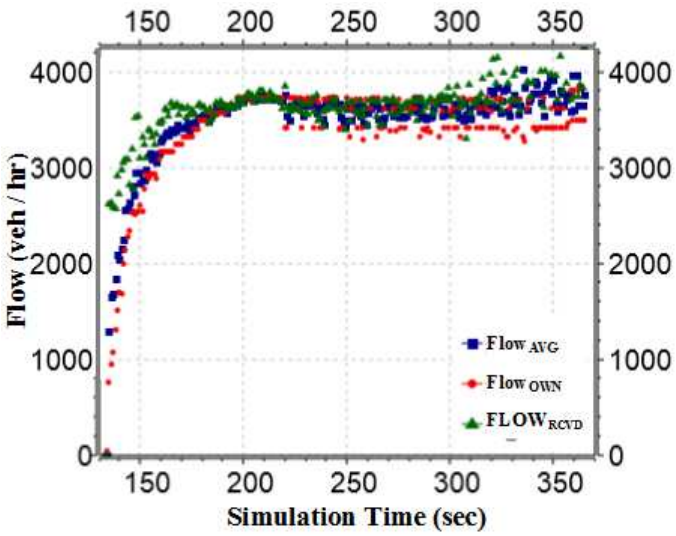


Fig. 9: Distribution of $Flow_{AVG}$, $Flow_{OWN}$ & $Flow_{RCVD}$ in case of Normal Traffic / No-Accident and all Honest Nodes

IDS in Fig. 10, when there are 20% rogue nodes. As shown in Figs. 10b the flow value goes down at first while the IDS runs the hypothesis tests to evaluate the received data and then starts to reject the false values. However, in the absence of the IDS (Fig. 10a) the Flow value is reduced as all the values are accepted.

4) Accident Scenario - Rogue Nodes: An accident scenario is simulated where rogue nodes start transmitting false (high) values after t=230 sec after an accident has occurred to deny the accident. The simulation is run both with and without the IDS and the results are shown in Fig. 11 (a) & (b) respectively. It can be seen in Fig. 11 (b) that the very high values by rogue nodes are being rejected by the IDS.

C. Evaluation Metrics

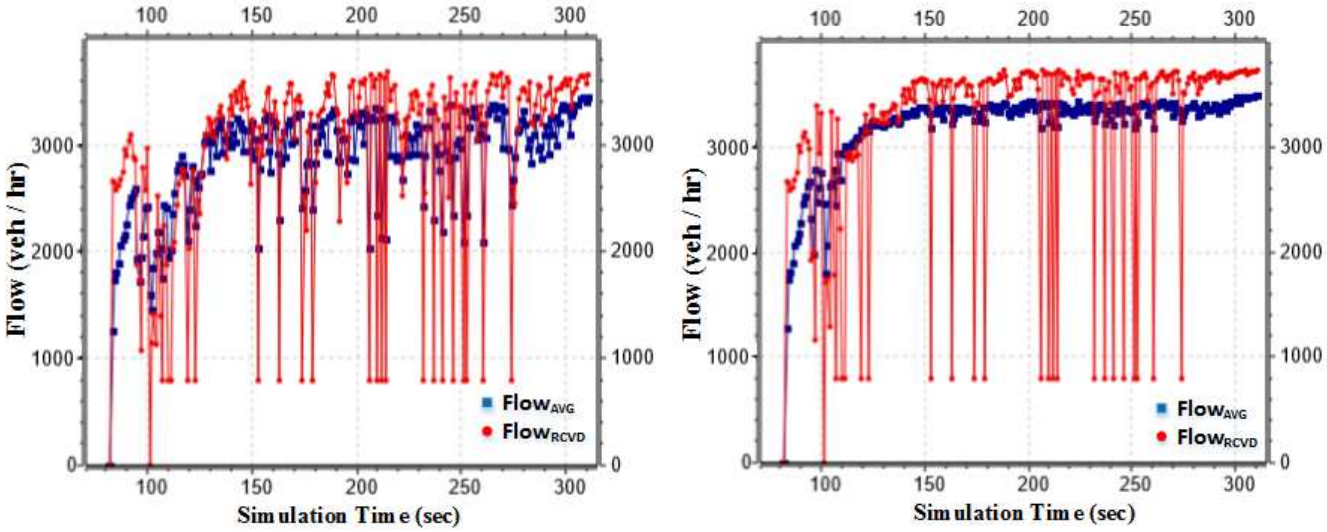
We test our IDS by computing the True Positive (TP) rate (detection rate), the false positive rate and the detection time. The number of rogue nodes was increased from 5% to 40% to test how successfully the proposed IDS classifies rogue nodes as rogue and honest nodes as honest. We also compare our results with that of two previous schemes that deal with false information attacks i.e. [15] and [24]. The metrics used are described below:

1. True Positive (TP): This is the detection rate of rogue nodes (RNs) i.e. what percentage of rogue nodes is detected and classified as rogue nodes. This is also referred to as sensitivity and is calculated as:

$$TP = \frac{No.\ of\ RNs\ detected\ correctly}{Total\ No.\ of\ RNs} \tag{7}$$

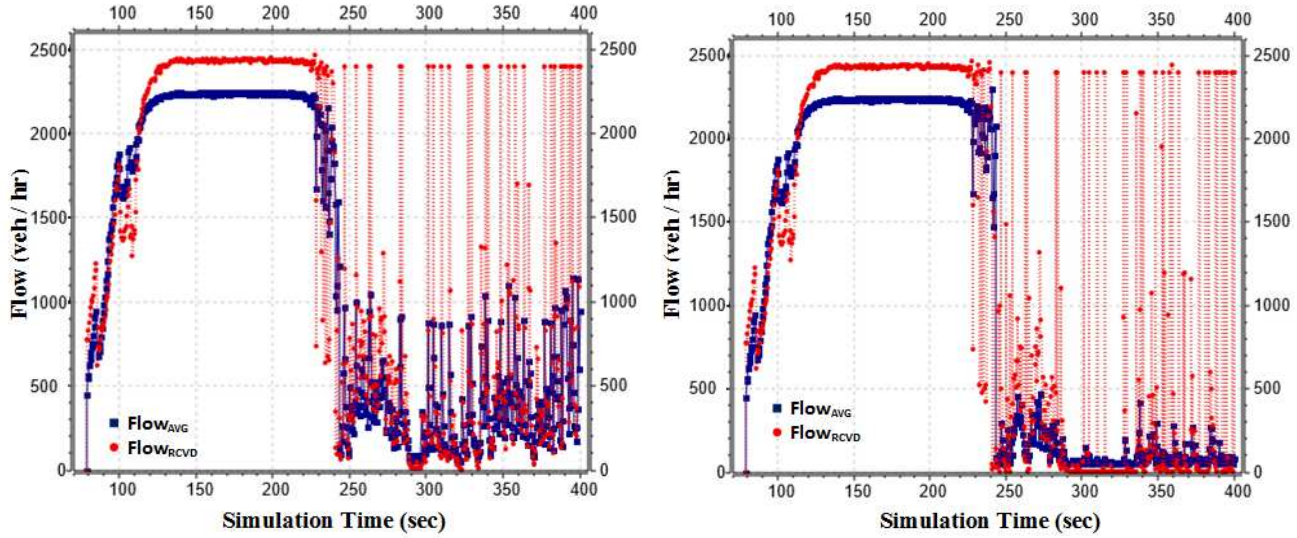
2. False Positive (FP): This is the percentage of honest nodes (HNs) incorrectly classified as rogue nodes. Specificity is defined as the number of honest nodes correctly identified and given as:

$$Specificity = \frac{No.\ of\ HNs\ identified\ correctly}{Total\ No.\ of\ HNs} \tag{8}$$



(a) $Flow_{AVG}$ Without IDS (b) $Flow_{AVG}$ WITH IDS

Fig. 10: No Accident Scenario: 20% Rogue Nodes - start transmitting false values at t=160sec



(a) $Flow_{AVG}$ Without IDS (b) $Flow_{AVG}$ WITH IDS

Fig. 11: Accident Scenario: 20% Rogue Nodes - start transmitting false values at t=230sec

and the false positives are calculated as:

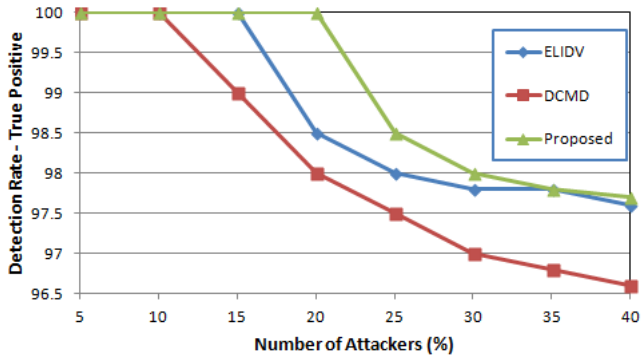
$$FP = 1 - Specificity \tag{9}$$

3. Overhead: The overhead is the cost incurred due to the IDS working and the extra data that is exchanged with other vehicles. It is an important metric as it is a measure of the efficiency of any scheme.

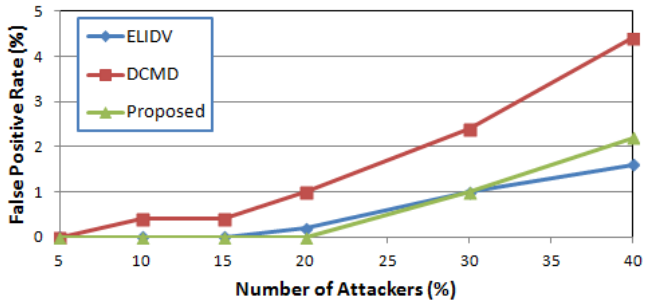
D. Effectiveness of Hypothesis Testing

The adoption of hypothesis testing works very well to determine whether the received data is correct or not. The t-test works very well to determine whether the data is false or not and thereby concluding that the node is rogue or honest. The t-test is comparing the population means of two populations and ascertaining if the means of the two populations are increasing or decreasing together. The simulation confirms that when the nodes are honest then the vehicles that are close together will have very similar Flow values (Fig. 9). This is true in all cases i.e. both in case of accident and free flowing traffic.

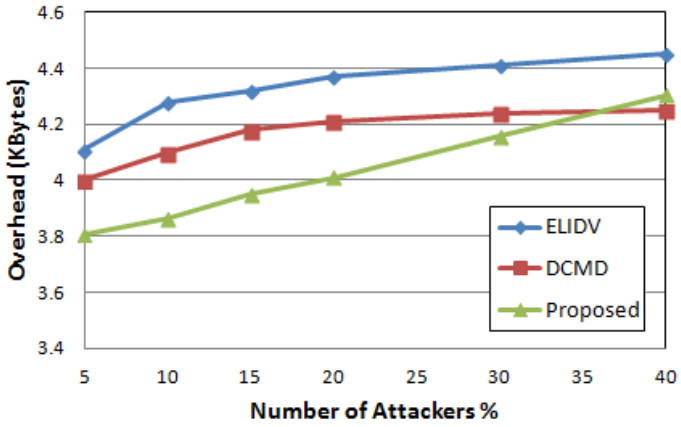
The cases simulated in this work are the worst case scenarios i.e. coordinated attacks by rogue nodes. This means that all rogue nodes work together and launch the attack at the same time to cause maximum damage. Such a coordinated attack is not only difficult to launch but also very expensive as it requires rogue vehicles to be placed together in strategic positions.



(a) Detection Rate Comparison in case of False Information Attack



(b) False Positive Rate Comparison in case of False Information Attack



(c) Overhead Comparison in case of False Information Attack

Fig. 12: Comparison of Proposed IDS: a)Detection Rate, (b) False Positive Rate, (c) Overhead

VI. DISCUSSION

In this section we discuss the performance of the proposed intrusion detection mechanism on the network, its reliability and robustness under changing parameters. We also compare the work to previously proposed approaches.

A. False Information Attack Detection

The proposed IDS is able to detect false information attacks very effectively by only analysing the data without taking into account any Trust or Reputation scores. The proposed mechanism is compared with two schemes i.e. DCMD [15] and ELIDV [24]. The detection rates are shown in Fig. 12a and false positive rates are compared in Fig. 12b. The detection rate (True Positives) of the proposed scheme is better than DCMD and ELIDV upto 30% rogue nodes and almost the same as ELIDV after that till 40%. The false positive rate of the proposed scheme is better than DCMD and ELIDV upto 20% rogue nodes but increases slightly above ELIDV at 40%.

B. Resilience to Sybil Attacks

In a Sybil attack, an attacker presents multiple identities with an intent to either create the illusion of congestion or accidents or deny their existence. So, a rogue vehicle will send multiple messages in order to cause confusion in the network by bringing the parameter value down. However, the proposed IDS aggregates the parameter values, therefore, the IDS will work very well and will be resilient to Sybil attacks as long as the total number of Sybil identities are less than 40% of the total identities (nodes) as shown in Fig. 12(a,b).

C. Overhead Comparison

The overhead of the proposed IDS is compared with the schemes in [24] and [15] and result is shown in Fig. 12c. The overhead in the proposed IDS is less as compared to DCMD and ELIDV except when there are 40% nodes at which point it is slightly higher than DCMD. The overhead in the proposed IDS increases with the increase in number of rogue nodes as the IDS starts to collect more past values to run the hypothesis test. However, the proposed IDS does not need to keep past parameter values as long as they agree with the calculated values which is the reason why the initial overhead is low.

D. Quick Response of IDS

The analysis shows that the test can be successfully conducted by taking only 7 samples from a rogue node i.e. the node that is incorrectly transmitting a false value, and performing the t-test on the population mean of two populations. The 7 samples can be collected in a minimum of 0.7 seconds if the beaconing rate is 100 ms. This means that the IDS enables the nodes to quickly decide whether to accept or reject the data received without generating a lot of overhead.

E. Countermeasures & Fault Tolerance

The proposed VANET model and exchange of parameters give the vehicular network a built-in resilience to launch countermeasures against false information attacks. The data is highlighted as false or malicious if it does not conform to the VANET model or if it fails the hypothesis test. The countermeasures include rejecting the data of that node and reporting the node as malicious. This was shown in Figs. 10b & 11b where the values were too low or too high as compared to the node's own values and were detected (and then rejected) by the IDS. The IDS is therefore, fault tolerant as it can work in the presence of false information.

F. Effective Information Dissemination

The widely proposed method of propagating emergency messages is by repeatedly broadcasting the message by vehicles to others behind them. This can quickly cause a broadcast storm in an already bandwidth limited channel. In the proposed scheme there is no channel congestion as there is no need for multi-hop retransmissions and the information is still disseminated effectively.

G. Limitations of proposed IDS

The proposed IDS works extremely well when the difference between the received values and the calculated values is high i.e. the values being received from the rogue nodes are too high or too low. However, if the rogue nodes coordinate and gradually decrease (or increase) their parameter values and launch the attack over some time then it will be very difficult to detect the attack. The reason is that the gradual decrease in the parameter values will not be flagged as an anomaly and thus never tested for correctness. However, as discussed previously doing this defeats the main purpose of the rogue / malicious vehicles i.e. to cause maximum damage or confusion in the network.

VII. CONCLUSION AND FUTURE WORK

In this work an intrusion detection system has been developed, tested and the results discussed. The results show that the proposed IDS is scalable and has an excellent performance when the number of rogue nodes is small. The performance degrades when the number of rogue nodes increases but still works reasonably well. The proposed model and IDS demonstrate the effectiveness of the statistical technique used to determine if the data is false based on the overall collected data without using Trust or reputation scores. The IDS does not depend on any infrastructure which is a major benefit as compared to other schemes. The false data is much easier to detect if it differs too greatly from the calculated data and difficult to detect if it varies slightly. However, the target of the rogue node is to drop or raise the value of its parameters quickly to damage the network and raising or dropping it gradually is not in its interest.

In the future, the work can be extended by modifying the IDS to detect other types of attacks in VANETs such as Denial of Service and false position reporting by rogue nodes in the network or a stationary user outside the network. This can be done by simulating the attacks using the developed platform and then detecting them with the help of anomaly or rule-based detection.

REFERENCES

[1] H. Sedjelmaci and S. M. Senouci, "A New Intrusion Detection Framework for Vehicular Networks," in *2014 IEEE Int. Conf. on Commun. (ICC)*, pp. 538-543. IEEE.

[2] H. Sedjelmaci, S.M. Senouci and M. Feham. "An efficient intrusion detection framework in cluster-based wireless sensor networks," *Security and Communication Networks*, 2013, pp. 1211-1224.

[3] A. Studer, E. Shi, F. Bai, and A. Perrig. "TACKing Together Efficient Authentication, Revocation, and Privacy in VANETs," *Proc. 6th Annual IEEE Commun. Society Conf. (SECON '09)*., pp. 1–9, 2009.

[4] D. Huang, S. Misra, G. Xue, and M. Verma, "PACP: An Efficient Pseudonymous Authentication Based Conditional Privacy Protocol for VANETS, *IEEE Trans. Intell. Transp. Syst*., vol. 12, no. 3, pp. 736-746, Sep. 2011.

[5] A.-N. Shen, S. Guo, D. Zeng, and G. Mohsen. "A Lightweight Privacy-Preserving Protocol using Chameleon Hashing for Secure Vehicular Communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, pp. 2543–2548, 2012.

[6] R. Lu, X. Li, T. H. Luan, X. Liang, and X. Shen. "Pseudonym Changing at Social Spots: An Effective Strategy for Location Privacy in VANETs," *IEEE Trans. Veh. Technol*., vol. 61, no. 1, pp. 86–96, 2012.

[7] M. Raya, and J-P. Hubaux, "Securing Vehicular Adhoc Networks," *Journal of Computer Security*, Special Issue on Security of Ad Hoc and Sensor Networks, vol. 15, no. 1, pp. 39–68, 2007.

[8] P. Papadimitratos, A. Kung, J. P. Hubaux, and F. Kargl, "Privacy and Identity Management for Vehicular Communication Systems: A Position Paper," in *Workshop on Standards for Privacy in User-Centric Identity Management*, Zurich, Switzerland, July 2006.

[9] R. Lu, X. Lin, H. Zhu, P. H. Ho, and X. Shen, "ECPP: Efficient Conditional Privacy Preservation Protocol for Secure Vehicular Communications", in *Proc. IEEE 27th Conf. Comp. Commun.* pp. 1229–1237, Apr. 2008.

[10] X. Lin, X. Sun, P.-H. Ho, and X. Shen, "GSIS: A Secure and Privacy Preserving Protocol for Vehicular Communication," *IEEE Trans. Veh. Technol*., vol. 56, no. 6, pp. 3442-3456, Nov. 2007.

[11] Vehicle Safety Communications Project Final Report: Identify Intelligent Vehicle Safety Applications Enabled by DSRC. s.l. :U.S. Department of Transportation, National Highway Traffic Safety Administration, 2005.

[12] K.A. Hafeez, L. Zhao, B. Ma, J.W. Mark, "Performance Analysis and Enhancement of the DSRC for VANET's Safety Applications," *IEEE Trans. Veh. Technol*., vol.62, no.7, pp.3069-3083, Sept. 2013.

[13] F. Kargl, P. Papadimitratos, L. Buttyan, M. Muter, E. Schoch, B. Wiedersheim, Ta-Vinh Thong, G. Calandriello, A. Held, A. Kung, J-P Hubaux, "Secure vehicular communication systems: implementation, performance, and research challenges," *IEEE Commun. Mag*., vol.46, no.11, pp.110-118, November 2008

[14] P. Golle, D. Greene, J. Staddon. "Detecting and correcting malicious data in VANETs," in *Proc. 1st ACM Int. workshop Veh. ad hoc networks*, pp. 29-37. ACM, 2004.

[15] S. Ruj, M. A. Cavenaghi, Z. Huang, A. Nayak, I. Stojmenovic. "On data-centric misbehavior detection in VANETs," in *Veh. Technol. Conf. (VTC Fall)*, 2011 IEEE, pp. 1-5. IEEE, 2011.

[16] M. Raya, P. Papadimitratos, I. Aad, D. Jungels, J.-P. Hubaux, "Eviction of misbehaving and faulty nodes in vehicular networks," *IEEE J. Sel. Areas Commun*, vol. 25, no. 8, pp. 1557-1568, Oct. 2007.

[17] P. Wex, J. Breuer, A. Held, T. Leinmuller, L. Delgrossi. "Trust Issues for Vehicular Ad Hoc Networks," *Veh. Technol. Conf. (VTC Spring 2008)*, pp.2800-2804, 11-14 May 2008.

[18] L. Qin, A. Malip, K. M. Martin, S. Ng, and J. Zhang. "A reputation-based announcement scheme for VANETs," *IEEE Trans. Veh. Technol.,*vol 61,pp 4095-4108,Nov. 2012.

[19] U. Minhas, J. Zhang, T. Tran, and R. Cohen, "Towards expanded trust management for agents in vehicular ad hoc networks, *Int. J. Comput. Intell. Theory Pract*., vol. 5, no. 1, pp. 3-15, Jun. 2010.

[20] A. Patwardhan, A. Joshi, T. Finin, and Y. Yesha, "A data intensive reputation management scheme for vehicular ad hoc networks," in *Proc. 3rd Annual Int. Conf. Mobile Ubiquitous Systems*, pp. 1-8. 2006.

[21] M. Baguena, S. Tornell, A. Torres, C. Calafate, J.-C. Cano, and P. Manzoni, "Vacamobil: Vanet car mobility manager for omnet++," in *2013 IEEE Int. Conf. on Commun. Workshops (ICC)*, pp. 1057-1061, Jun. 2013.

[22] M. Behrisch, L. Bieker, J. Erdmann, and D. Krajzewicz. "Sumo - Simulation of Urban Mobility: An overview," in *SIMUL 2011, 3rd Int. Conf. on Advances in System Simulation*, pages 63-68, Oct. 2011.
[23] K. Zaidi, M. Milojevic, V. Rakocevic, and M. Rajarajan. "Data-centric Rogue Node Detection in VANETs," in *2014 IEEE 13th Int. Conference on Trust, Security and Privacy in Computing and Commun. (TrustCom)*, pp. 398-405.
[24] H. Sedjelmaci, S.M. Senouci, M.A. Abu-Rgheff. "An Efficient and Lightweight Intrusion Detection Mechanism for Service-Oriented Vehicular Networks," *IEEE Internet things J.*, vol. 1, No.6, Dec 2014.
[25] N. Bimeyer, C. Stresing, and K.M. Bayarou. "Intrusion detection in VANETs Through Verification of Vehicle Movement Data," in *Vehicular Networking Conference (VNC)*, 2010 IEEE.
[26] E.M. Shakshuki, N. Kang and T. R. Sheltami. "AACKA Secure Intrusion-Detection System for MANETs," *IEEE Trans. on Ind. Electron.*. 60(3), pp. 1089-1098. 2013.
[27] J. Hortelano, J.C. Ruiz, P. Manzoni. "Evaluating the Usefulness of Watchdogs for Intrusion Detection in VANETs," in *IEEE Int. Conf. Commun. Workshops (ICC)*, 2010
[28] H. Sedjelmaci, T. Bouali, S.M. Senouci. "Detection and Prevention From Misbehaving Intruders in Vehicular Networks," in *IEEE GLOBECOM 2014*, Austin, USA, 8-12 December, 2014.
[29] T. Gazdar, A. Rachedi, A. Benslimane, A. Belghith. "A Distributed Advanced Analytical Trust Model for VANETs," in *IEEE GLOBECOM*, California, USA, 2012, pp. 201-206.
[30] H. Sedjelmaci, S.M. Senouci. "An accurate and efficient collaborative intrusion detection framework to secure vehicular networks," in *Computers & Electrical Engineering* 43 (2015): 33-47.

Kamran Zaidi received the B.E degree in Electrical Engineering from National University of Sciences & Technology (NUST), Pakistan in 1999 and his MSc in electronics engineering from London Metropolitan University, UK in 2003. He is currently pursuing his PhD degree in Information Engineering at City University London. His research interests include security and privacy of wireless and wired networks with focus on intrusion detection, cryptography and identity management.

Milos B. Milojevic received the BSc and MSc degrees in electrical engineering from the School of Electrical Engineering, University of Belgrade, Serbia, in 2009 and 2010. He is working towards his PhD degree in Electronic Engineering at City University London that he started in 2011. His research interests include vehicular ad-hoc networks, intelligent transport systems, data aggregation and message dissemination in vehicular ad-hoc networks.

Veselin Rakocevic (M'01) received his Ph.D. degree from Queen Mary, University of London, United Kingdom, in 2002, and his Dipl.Ing. degree from the University of Belgrade, Serbia, in 1998, both in Electronic Engineering. He currently works as Reader in Electronic Engineering at City University London, UK, where he has been since 2002. His main research interest is in the operation of multihop wireless networks, especially addressing the problems of optimal scheduling, rate control, and quality of service, with application in vehicular networks and wireless sensor networks.

Arumugam Nallanathan (S'97-M'00- SM'05) is a Professor of Wireless Communications in the Department of Informatics at King's College London (University of London). He served as the Head of Graduate Studies in the School of Natural and Mathematical Sciences at King's College London, 2011/12. He was an Assistant Professor in the Department of Electrical and Computer Engineering, National University of Singapore from August 2000 to December 2007. His research interests include 5G Technologies, Millimeter wave communications, Cognitive Radio and Relay Networks. In these areas, he co-authored more than 250 papers. He is a co-recipient of the Best Paper Award presented at the 2007 IEEE International Conference on Ultra-Wideband (ICUWB2007). He is a Distinguished Lecturer of IEEE Vehicular Technology Society.

He is an Editor for IEEE Transactions on Communications and IEEE Transactions on Vehicular Technology and a Guest Editor for IEEE Transactions on Emerging Topics in Computing: Special Issue on Advances in Mobile and Cloud Computing. He was an Editor for IEEE Transactions on Wireless Communications (2006-2011), IEEE Wireless Communications Letters and IEEE Signal Processing Letters. He served as the Chair for the Signal Processing and Communication Electronics Technical Committee of IEEE Communications Society, Technical Program Co-Chair (MAC track) for IEEE WCNC 2014, Co-Chair for the IEEE GLOBECOM 2013 (Communications Theory Symposium), Co-Chair for the IEEE ICC 2012 (Signal Processing for Communications Symposium), Co-Chair for the IEEE GLOBECOM 2011 (Signal Processing for Communications Symposium), Technical Program Co-Chair for the IEEE International Conference on UWB 2011 (IEEE ICUWB 2011), Co-Chair for the IEEE ICC 2009 (Wireless Communications Symposium), Co-chair for the IEEE GLOBECOM 2008 (Signal Processing for Communications Symposium) and General Track Chair for IEEE VTC 2008. He received the IEEE Communications Society SPCE outstanding service award 2012 and IEEE Communications Society RCC outstanding service award 2014.

Muttukrishnan Rajarajan (Raj) is a full-professor of Security Engineering at City University London, United Kingdom. He leads the information security group at City and his research interests are in the areas of identity, privacy and intrusion detection. He has published more than 200 papers in these areas and continues to be involved in the editorial boards and technical programme committee of several international conferences and journals. He has actively participated in several cyber security debates in UK, Europe and internationally and continues to act as an advisor to the Government of India cyber security labs in the area of SCADA security and identity management. Raj is an active member of the Institute of Information Security Professionals, UK and advises the UK Government's identity assurance programme in the areas of access control and privacy. Raj was chosen as one of the leading academics with outstanding research impact in the security community. He is a Senior Member of IEEE and an advisory member of the Institute of Information Security Professionals, United Kingdom.