International Journal of Computer Networks and Applications (IJCNA)
DOI: 10.22247/ijcna/2021/208890 Volume 8, Issue 2, March – April (2021)

RESEARCH ARTICLE

Ensemble Miscellaneous Classifiers Based Misbehavior Detection Model for Vehicular Ad-Hoc Network Security

S. Sumithra
Department of Information Technology, Bharathiar University, Coimbatore, Tamil Nadu, India.
sumiphdit@gmail.com

R. Vadivel
Department of Information Technology, Bharathiar University, Coimbatore, Tamil Nadu, India.
vlr_vadivel@yahoo.co.in

Received: 12 February 2021 / Revised: 02 March 2021 / Accepted: 10 March 2021 / Published: 27 April 2021

Abstract – Vehicular Ad-Hoc Network is an emerging technology, mainly developed for road safety applications, entertainment applications, and effective traffic conditions. VANET applications work based on the accurate mobility information shared among the vehicles. Sometimes attackers manipulate the mobility information shared by the adjacent vehicle or neighboring vehicle, which results in terrible consequences. To deal with the illusion-based type of attacks, researchers have proposed enormous solutions. Unfortunately, those solutions could not deal with the dynamic vehicle conditions and variable cyber malfunctions, which reduces the misbehavior detection accurateness and increases the false-positive rate. In this paper, the dynamic vehicle context is taken into account to propose a two solutions such as Miscellaneous VANET Classifiers based Misbehavior Detection Model (MVC-MDM) and Ensemble Miscellaneous VANET Classifiers based Misbehavior Detection Model (EMVC-MDM). This model is constructed based on the Mobility Data Gathering phase, Mobility Context Feature Extraction phase, Mobility Context Feature Level Fixing phase, Hampel Filter based Context Reference Building phase, Constructing Miscellaneous VANET Classifiers based Misbehavior Detection model and Ensemble Miscellaneous VANET Classifiers based Misbehavior Detection phase. Vehicle context is prepared using the data-centric features and the behavior-based features of the vehicles. The Nonparametric Hampel filter and Kalman filter are used to building the context reference model. These filters discover the temporal and spatial correlation of the uniformity in the current mobility information. Vehicle features are extracted locally according to the stability, likelihood, and performance of the vehicles' mobility information. A random forest based learning algorithm is used to train and test the classifiers. The proposed MVC-MDM and EMVC-MDM has been simulated in various context scenarios and the presence of misbehaving vehicles. NGSIM dataset has been used for extensive simulation. The results prove that the effectiveness and the reliability of the proposed MVC-MDM and EMVC-MDM are higher than the existing misbehavior detection systems.

Index Terms – Stability, Likelihood, Performance, Hampel Filter, Kalman Filter.

1. INTRODUCTION

In Vehicular Ad-hoc Network, Vehicle automation has become true due to the advancements of artificial intelligence in embedded systems. Vehicles in the VANET environment gather their position information from the On-Board Units (OBU) such as sensors like GPS, accelerometer, speedometer, cameras, RFID, etc. [1]. The vehicles communicate with each other which is Vehicle to Vehicle (V2V) or with the Road Side Units which is the Vehicle to Infrastructure (V2I) use Wireless Access for Vehicular Environment (WAVE) technology. By communicating with each other, the vehicles can extend their observation beyond their OBU. So that vehicles can be aware of unknown attacks and false messages up to their communication range of *1km*.

In the VANET environment, automated vehicles get the ability to find dangerous driving situations such as accidents, damaged roads, and bad weather conditions. VANET applications heavily depend upon the accuracy and the reliability of the mobility information shared by other adjacent vehicles (neighboring vehicles). Mobility information sends by the adjacent vehicles contains the details of the adjacent vehicle such as position, velocity, direction, acceleration, driving status, etc. This mobility information is packed as a beacon message and broadcast to other vehicles. The purpose of this mobility message is to warn other adjacent vehicles about the hazards during their travel [2]. The VANET applications fixed in the adjacent vehicles use the mobility information to predict the true position of the broadcasting vehicle. This prediction improves road safety and enhances network performance. Basic Safety Messages (BSM) and the Cooperative Awareness Messages (CAM) are the fundamental safety messages (BSM.1 IEEE standard and CAM ETSI standard). Both the safety message standards ensure that the vehicles share their mobility information to all the vehicles crossing their communication range at a high rate within a single-hop communication. In VANET, vehicles should broadcast 10 messages per second within a 1 km communication range. In such situations, mobility messages encounter attacks during the transmission and loss integrity. Therefore, VANET applications must have strong security models [3]. In this paper, Misbehavior Detection System and Misbehavior Detection Model represent the same meaning.

VANET applications subject to enormous types of cyber-attacks. When vehicles share their mobility information, attackers interrupt the message, manipulate their context, and spread false information. Therefore, the integrity of those mobility messages is very important. For Instance, malicious software can damage the vehicles' computing system and the attackers could easily control the vehicle. This type of vehicle is called misbehaving vehicles. Sometimes misbehaving vehicles share false information about the vehicle. False information may cause disastrous events such as accidents. Illusion attacks are common attacks in the Vehicular network, which create an imaginary position of the vehicle and trigger the adjacent vehicle to react to the unreal events [4]. Attackers steal the vehicle control and use that vehicle for misbehaving activities such as assassinations, terrorist activities, kidnapping, redirect the traffic flow to make accidents or create heavy traffic. Such misbehaving acts may force the vehicles to meet with critical situations. This situation results in significant changes in traffic flow, bandwidth utilization, and the performance of the VANET applications. Security challenges and the possible faced by VANET applications are discussed in [5] [6] [7]. Based on the studies, faulty data-based attacks or illusion attacks cause high threats and damages and they are difficult to detect. Detection problems are one of the main challenges in VANET. This research work is focused on the misbehavior detection process of vehicles that spread faulty data.

Based on cryptographic techniques, numerous solutions have been offered to increase the truthfulness of the mobility information of vehicles [8] [9] [10]. Unfortunately, these solutions are lacking in detecting the misbehaving up to a satisfactory level. Misbehaving vehicles continue to manipulate mobility data and spread false information among vehicles. Cryptographic techniques easily allow attackers to alter the mobility data before it starts its job. To safeguard from misbehaving vehicles, it is important to detect those vehicles and be aware of the faulty message send by those vehicles. Studies talk about many solutions for the problems in misbehavior detection. However, most of these existing solutions are unable to detect faulty data, which results in vulnerable situations. The related work section reviews the existing misbehavior detection models in detail. To solve misbehaving detection problems, a new Ensemble Miscellaneous VANET Classifiers based Misbehavior Detection Model (EMVC-MDM) is proposed in this paper. The proposed EMVC-MDM is constructed with six phases as follows, the mobility data gathering phase, mobility context feature extraction phase, mobility context feature levels, Hampel filter based context reference building, constructing Miscellaneous VANET Detection Model (EMVC-MDM). The Broadcasting Vehicle (BV) gathers the mobility information using the Kaman filter in the first phase. Mobility information contains the vehicles' position, velocity, direction, and acceleration. The uncertainty present in the mobility message is removed by the Optimal Innovation based Adaptive Estimation Kalman filter (OIAE-KF) algorithm. The next step is broadcasting the mobility information effectively to the adjacent vehicles (neighboring vehicles) using Traffic Condition Aware Customized Beacon Broadcasting (TCA-CBB) method.

The second phase is the mobility context feature extraction phase. In this phase, features are selected from the mobility information to classify the vehicles. The Stability based feature is selected to analyze the position prediction error and velocity prediction errors. The Performance based feature is selected to examine the vehicle behavior. The likelihood based feature is extracted to study the transmissions within the range of the vehicle and the intersection within the range of the vehicle. The third phase is setting up the levels of the extracted features. The stability level (SL) feature is set at Level 1 (L1) while the Likelihood Level (LL) feature is set at Level 2 (L2) for transmission range based feature and Level 3 (L3) for intersection based feature. The performance based feature is set at Level (L4). The fourth phase is Hampel filter based context reference building where; contextual references are built using the features extracted from the mobility information. Hampel filter is used to remove the outliers from the context reference.

The fifth phase constructs the Miscellaneous VANET Classifiers using the contextual references built from the features. The standard z- score outlier detection method is used to perform the reliability computation. The output of the standard Z-score determines whether the adjacent vehicle is a trusted vehicle or misbehaving vehicle. Each feature is tested for the standard z- score and constructed the classifiers. These classifiers are independent classifiers without taking the correlation between the classifiers. The sixth phase analyzes the correlation between the classifiers and creates an ensemble miscellaneous classifier. The random forest method is used for creating ensemble classifiers. Each classifier is trained using random sample features. The majority voting scheme combines the outputs of the classifiers and the misbehaving detection rate is improved by the ensemble approach. The outputs of the final random forest based ensemble classifiers are analyzed under the weighted average for the final decision to conclude whether the adjacent vehicle is a trusted or misbehaving vehicle.

The contribution of the paper is as follows:

Mobility based features are extracted from the stability, performance, and likelihood characteristics of a vehicle to effectively analyze the context. The mobility context feature level is determined for each derived feature.

Mobility Context reference is created by the Hampel filter, which removes the outliers using the temporal and spatial relationship of the mobility information. The Scatter plot method visualizes the data observed.

Miscellaneous VANET Classifiers based Misbehavior Detection Model (MVC-MDM) is built based on the Hampel filter based context references. These classifiers perform misbehavior detection independently on each feature. The standard Z-score algorithm is used to enhance the misbehavior detection accuracy and covers enormous attacks.

Finally, all the classifiers are integrated to form an Ensemble Miscellaneous Classifiers based Misbehavior Detection Model (EMVC-MDM). Without ensemble, the classifiers work independently and not considering the correlation between the classifiers. This approach is vulnerable to VANET applications. Therefore, an ensemble based miscellaneous classifier is constructed. The output of the Miscellaneous VANET classifiers (MVC-MDM) is given as the input to the Ensemble Miscellaneous Classifiers based Misbehavior Detection Model (EMVC-MDM) which effectively detects the misbehaving vehicles.

This paper is organized as an introduction in section 1and related works are discussed in section 2. Section 3 explains the proposed method EMVC-MDM. Performance evaluation of the proposed method is discussed in section 4. A detailed description of the results obtained by the simulation of the proposed method is described in section 5. Conclusion and reference of future research are discussed in section 6.

2. RELATED WORK

Several solutions have been proposed for misbehavior detection. Misbehavior Detection System (MDS) is one of the important studies in VANET. Misbehavior detection is classified based on three approaches such as behavioral based, data centric based and hybrid based approaches. A complete study if these approaches are analyzed in reference [11]. Behavior based approach is defined as if a vehicle does not act according to the designed behavior of the VANET protocols and applications. Watchdog mechanism is the popular method used for behavioral based attacks [12] which frequently monitors the vehicles' behavior such as broadcasting behavior and forwarding behavior in the VANET routing protocol. For instance, observing the broadcasting behavior helps to detect the Denial of Service (DOS) attacks, and observing the forwarding behavior help in detecting wormhole attacks. Even though there are some limitations in behavioral based detection. It is not suitable for vehicles, which send defective data; moreover, behavioral based detection only focuses only on observing the behavior against the protocols. Unfortunately, vehicles, which spread faulty data, will neither go against the protocol rules nor act against the expected behavior. Another approach is data centric based misbehavior detection [13]. The Data centric approach is again categorized as context based or event based misbehavior detection. False congestion warnings, fake crash notifications, and bogus emergency messages could be detected by the event based misbehavior detection system. Even though the event based detection model performs well, it is limited to only a certain type of events such as accidents, crashes, and congestion. But this type of approach is suggested by many authors. [14].

Context based data centric is an interesting approach where most of the VANET protocols and applications rely on the accurate context information. Therefore, the context based data centric approach has several advantages in misbehavior detection. Different types of attacks can be detected by this approach such as spreading fake event messages and bogus congestion information [15]. VANET attackers mostly concentrate on the context of the mobility information, so that they could easily create extremely dangerous situations while attackers remain undetected. This paper focuses on the mobility information send by the broadcasting vehicle to develop the misbehavior detection model. This type of MDS is more feasible than the event based MDS. Because event based MDS attempts to detect the attack at the early stage [16].

Data centric based MDS use likelihood checking and stability checking in the mobility information. Likelihood based detection makes use of the real time data for detecting, unlike data. For instance, a vehicle cannot exist in various locations at the same time. Stability based MDS compares the origin of the message coming from various senders. For instance, the speed measured by the speedometer should be the same as the velocity generated by the positioning sensors of the vehicle. Under controlled conditions, stability and likelihood based approaches could detect the attacks effectively and the reliability and the accuracy of the mobility information can be gathered [17].

However, VANET applications are highly dynamic, and controlled conditions will not be appropriate. Besides, the accuracy and trustworthiness of the mobility information are doubtful in these controlled conditions. All the existing misbehavior detection systems use static threshold values to identify the misbehavior. This static threshold based approach results in a low detection ratio and high false positive rate.

Authors of Data Centric Context Aware Misbehavior Detection System (DCA-MDS), proposed an effective MDS, which moderately increased the detection accuracy and reduced the false positive rate. Unfortunately, this approach was unable to detect the misbehaving vehicle. DCA-MDS is exposed to VANET context based attackers who manipulate the context of the mobility information [18].

To overcome this issue, the only solution is to integrate the data –centric based approach with a behavioral based approach. In this paper, the above said integration approach is called the Miscellaneous based approach. The theory is that the context based attacker will be interested in injecting false information into the victim's vehicle. For instance, incrementing broadcast rate will result in unusual behavior of the vehicles which meet the same context compared to adjacent vehicle.

Bissmeyer and Michal proposed a model that combines data-centric based detection and behavioral based detection, which is named Entity Centric Trust based Misbehavior Detection System (ECT-MDS) [16]. Ghaleb et al proposed the Misbehavior-Aware On-Demand Collaborative Intrusion Detection System (MA-CIDS) based on the ensemble learning method to enhance the efficiency of intrusion detection in VANET [17]. The outputs of the classifiers are used as the input for the robust weighted voting scheme. But the authors have used the static thresholds for detection. Moreover, the classifiers are used as independent classifiers without considering the correlation among them. Thus Bissmeyer model is exposed to attacks such as an imaginary situation where the attackers spread imaginary traffic situations. Stubing’s MDS (SMDM) combines data centric and behavioral based techniques [18].

SMDS is the baseline for comparison which was used by many researchers. To defend against an extensive range of misbehaviors and attackers, data-centric and behavioral based miscellaneous misbehavior detection model is a suitable approach. Many researchers in related works have suggested this approach. Most of the existing misbehavior detection models are based on static thresholds, which is not suitable for an ephemeral network like VANET.

In most cases, the correlations between the classifiers were not considered. When the classifiers act independently, they could not produce effective results. Therefore, ensemble based learning approaches such as Random forest, which combines the classifiers, and generates appropriate results. Zhang and Chunhua proposed Dempster – Shafer theory (DST) and Support Vector Machine-based misbehavior detection mechanism as the solution for false message attack. This mechanism works based on the data trust and the vehicle trust. Even though the DST mechanism does not consider the correlation between the classifier [19].

3. THE PROPOSED ENSEMBLE MISCELLANEOUS VANET CLASSIFIERS BASED MISBEHAVIOR DETECTION MODEL (EMVC-MDM)

To resolve the problem of illusion-based attacks in VANET, the proposed Ensemble Miscellaneous Classifiers based Misbehavior Detection Model (EMVC-MDM) constructs multifaceted classifiers based on the features extracted from the mobility information. Unsupervised non-parametric methods such as Hampel filter and Kalman filter are used for context representation. Mobility information is the backbone for VANET applications, which is the target for many attackers. For instance: Illusion based attacks that create an imaginary situation or spreads a false message among the vehicles. The proposed method is constructed of a set of miscellaneous unsupervised classifiers to filter the outliers, which enhances the efficiency of the misbehavior detection model.

The proposed EMVC-MDM is deployed in each connected vehicle to identify the misbehavior at the very early stage before starting the attack. Therefore, the proposed EMVC-MDM is a host-based model. The features of the mobility information are used for composing the context reference, which trains the ensemble classifiers. Unsupervised statistical methods such as the Hampel filter is used so that the model could adapt to any alteration in the context. Kalman filter first gathers the recent mobility data from the adjacent vehicle. The whisker plot is a method to summarize the mobility information on a regular interval, which shows the shape of the distribution and the variability. Hampel filter identifies the outliers in the distribution. The outliers of the Kalman filter, Whisker plot, and Hampel filter are considered as misbehaviors.

The proposed EMVC-MDM is developed based on six phases. The mobility data gathering phase, mobility context feature extraction phase, mobility Context features levels, Hampel filter based context reference building phase, constructing miscellaneous VANET classifiers, and Ensemble miscellaneous VANET classifiers based misbehavior detection model.

3.1. Mobility Data Gathering Phase

The Mobility Data Gathering phase is made up of two main units. Obtaining mobility information from the individual vehicle and broadcasting the mobility data among the adjacent vehicles. The adjacent vehicle can recreate the absent data in its surrounding area. After obtaining the mobility information from the adjacent vehicle, the data is prepared for context representation.

3.1.1. Obtaining Mobility Information

The positioning sensors such as gyro sensor, speedometer, and accelerometer are used to obtain the mobility data from each vehicle. Unfortunately, these sensors are vulnerable to environmental noise and dynamic traffic conditions. This issue is treated with an enhanced Kalman Filter technique, which is also called Optimal Innovation based Adaptive Estimation Kalman Filter (OIAE-KF) [20]. The uncertainty encountered in the observations of positioning sensors is rectified using OIAE-KF. The main benefit of OIAE-KF is to estimate the measurement noise covariance and automatically swap the observations from Adaptive Kalman Filter to Dead Reckoning. OIAE-KF improves the position prediction accuracy in dynamic noise environments.

3.1.2. Broadcasting the Mobility Information

A high broadcasting rate is not suitable for the VANET environment due to its ephemeral characteristics. Mobility information is accessed by all the vehicles, so the high broadcasting rate decreases the performance of the VANET application. Traffic Condition Aware Customized Beacon Broadcasting (TCA-CBB) method concentrates on the issue and moderately reduces the broadcasting rate according to the vehicle traffic condition. TCA-CBB is constructed based on two main components, namely Broadcasting Vehicle Segment (BVS) and Adjacent Vehicle Segment (AVS). BVS gathers the mobility information from the adjacent vehicle.

Traffics condition estimation unit calculates the model parameters of the mobility information in terms of position, velocity, and direction. Self-position estimation module predicts the recent position of the broadcasting vehicle. The error threshold determines the importance of the beacon message. If the importance of the beacon message is high, a complete beacon is constructed and broadcasted. The adjacent vehicle receives the broadcasted beacon message and predicts its current position and verifies it with the received position. The omitted or lost message is reconstructed at the adjacent vehicle. Figure 1 shows the diagrammatic representation of the mobility data gathering phase.

3.2. Mobility Context Feature Extraction Phase

When mobility data gathering from the adjacent vehicle is completed, context features are derived from the mobility data. Three main categories of features are derived, such as stability-based features, performance based features, and Likelihood based features.

3.2.1. Stability Based Features

The vector of the divergence between the estimated and the actual received mobility data, which is also called innovation error, is considered as the stability based features. Therefore, four measurements such as longitude position prediction error, latitude position prediction error, longitude, speed prediction error, and latitude, speed prediction error are considered as stability based features with fixed acceleration and two-dimensional state vector.

1. *X-axis Position Prediction Error* - The variation between the longitude measurements received from the adjacent vehicle and the predicted one by using the TCA-CBB algorithm.
2. *Y-axis Position Prediction Error* - The variation between the latitude measurements received from the adjacent vehicle and the predicted one by using the TCA-CBB algorithm.
3. *X-axis Velocity Prediction Error*- The ratio of the change of longitude speed prediction error
4. *Y-axis Velocity Prediction Error*- The ratio of the change of latitude speed prediction error

3.2.2. Likelihood Based Features

Two main characteristics are considered as the Likelihood Based Features, such as Transmission range based feature and Intersection based feature. Transmission Range Based Feature, which is the distance between all broadcasting vehicles and the receiver vehicle (holds the MDM). This distance is measured using a Euclidean distance calculation. Intersection Based Feature, which represents the number of times vehicle intersecting adjacent vehicle's communication area.

3.2.3. Performance Based Features

The following features, Broadcasting Rate, Broadcasting Delay, Jerk Acceleration, Received Beacons, Velocity deviation and connection length represent the vehicles' performance. These features are considered Performance Based Features.

1. *Broadcasting Rate*- The average of the number of successfully received beacons in a certain connection length.
2. *Broadcasting Delay*- The average difference between the beacon creation time and received time at the adjacent vehicle.
3. *Jerk Acceleration*- The ratio of the difference in acceleration.
4. *Received Beacons*- The total number of beacons received at the adjacent vehicle from the sender vehicle.
5. *Velocity deviation*- The difference between the velocity of the sender vehicle and the adjacent vehicles
6. *Beacon time epoch*- The time difference between the current epoch time and the first received beacon message from the adjacent vehicle.

Category	Features	
Stability Based Features	X-axis Position Prediction Error	SL (L1)
	Y-axis Position Prediction Error	
	X-axis Velocity Prediction Error	
	Y-axis Velocity Prediction Error	
Likelihood Based Features	Transmission Range Based Feature	PL(L2, L3)
	Intersection Based Feature	
Performance Based Features	Broadcasting Rate	LL(L4)
	Broadcasting Delay	
	Jerk Acceleration	
	Received Beacons	
	Velocity Deviation	
	Beacon Time Epoch	

Table 1 Mobility Context Features

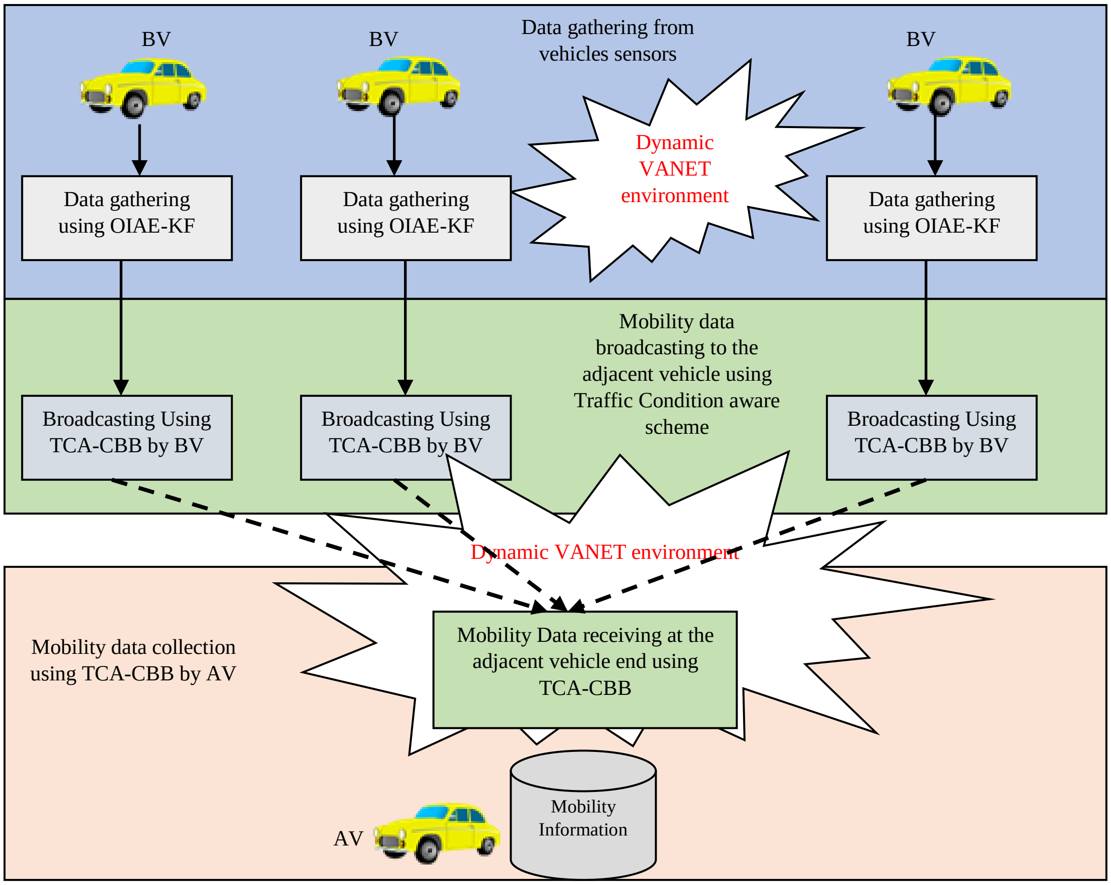

BV
BV
Data gathering from vehicles sensors
BV
Dynamic VANET environment
Data gathering using OIAE-KF
Data gathering using OIAE-KF
Data gathering using OIAE-KF
Mobility data broadcasting to the adjacent vehicle using Traffic Condition aware scheme
Broadcasting Using TCA-CBB by BV
Broadcasting Using TCA-CBB by BV
Broadcasting Using TCA-CBB by BV
Dynamic VANET environment
Mobility data collection using TCA-CBB by AV
Mobility Data receiving at the adjacent vehicle end using TCA-CBB
Mobility Information
AV

Figure 1 Mobility Data Gathering Phase

3.3. Mobility Context Feature Levels

Using the temporal and spatial findings between the adjacent vehicles' mobility information, the contextual references are created. Three main features are derived which constructs the mobility context, such as stability based features, Likelihood based features, performance based features. The Stability based features decide the Stability Level (SL) of each adjacent vehicle. Stability Level is represented as L1 in this paper. Likelihood based features decide the Likelihood Level LL) of each adjacent vehicle. Possibility Level (PL) is subdivided into Transmission Range Level (L2) and

Intersection Level (L3). Performance based features decide the Performance Level (PL) that examines the vehicles' communication behavior. L4 represents the Performance Level. The algorithm shows the systematic process of feature extraction mobility information to derive the mobility context reference. OIAE-KF and TCA-CBB are used to generate the mobility information and the Whisker Plot visualizes the data received. Table 1 illustrates the features and their levels derived from the mobility information.

3.3.1. Stability Level (SL) – L1

Each vehicle gathers the mobility information using mobility data gathering modules. Due to the ephemeral network and harsh surroundings in VANET, mobility data gathering sensors and positioning units such as GPS may contain uncertainty in measurement. To overcome this issue, an Optimal Kalman filter based algorithm (OIAE-KF) and Traffic condition aware broadcasting method is used. The Kalman filter is used to measure and predict the mobility data from an adjacent vehicle. The innovation sequence shows the variation between the actual mobility data and the measured mobility data. Innovation sequence is also used to construct the stability of the mobility information. Whisker plot mechanism visualizes the innovation sequence predicted at various time epochs to minimize the error between the innovation sequence send by the adjacent vehicle. The Stability based feature is extracted in two ways. A series of vectors represents the Innovation Sequence (IS) measured at each time epoch for each adjacent vehicle as shown in equation (1).

$$IS^{BV(i)} = \{\ldots\ldots z_{k-1}^{AV(i)}, z_{k}^{AV(i)}, z_{k+1}^{AV(i)}, \ldots\ldots\} \quad (1)$$

The following equation (2) shows how to calculate the innovation sequence error of an adjacent vehicle.

$$z_k^{AV(i)} = \begin{cases} y_k^{AV(i)} - \breve{y}_{k|k-\lambda(i)}^{AV(i)} & if\ beacon\ is\ successfull \\ z_{k-\lambda(i)}^{AV(i)} \times \left(k - \lambda(i)\right) & else \end{cases} \quad (2)$$

Where $z_k^{AV(i)}$ is the vector that holds the innovation error at the time epoch k of the adjacent vehicle AV(i). $y_k^{AV(i)}$is the vector that holds the mobility information received at the time epoch k from adjacent vehicle AV(i). $\breve{y}_{k|k-\lambda(i)}^{AV(i)}$is the vector that holds the mobility information predicted at time epoch k using last received beacon messages and adjacent vehicle prediction method (TCA-CBB). λ(i)represents the time epoch of the last received beacon. The next step is each vehicle visualizes the innovation errors and removes the outliers, which are denoted as $O_k^{AV(i)}$in each vehicle's locality. $O_k^{AV(i)}$is generated using the Scatter Plot method. The Scatter Plot is used for data visualization and shows the relation between them. Outliers are found due to the randomness of the innovation sequence and the harshness of the VANET environment. Therefore, the Stability Level (SL), which is L1, is used to estimate the mobility information stability of the vehicle at time epoch k. The stability Level could be calculated as follows where m is the slope of the graph, $M_{k(i)}$is the upper limit of the Scatter plot and $z_k^{AV(i)}$is the innovation error as shown in equation (3) and equation (4).

$$Scatter\ plot\ (y) = mx + b \quad (3)$$

$$L1_k^{AV(i)} = \max(M_{k(i)}, z_k^{AV(i)}) \quad (4)$$

3.3.2. Likelihood Level (LL) – L2, L3

Likelihood based features decide the Likelihood Level (LL) of each adjacent vehicle. Likelihood Level (PL) is subdivided into Transmission Range Level (L2) and Intersection Level (L3).

1. Transmission Range based Likelihood Level - L2

It is a fact that vehicles in the VANET environment can communicate and exchange data only within a certain range. Based on this fact transmission range based feature is derived. As mentioned earlier, due to the dynamic nature and harsh environment, the communication range may vary in context. Thus the new vehicles' communication range is selected as a transmission range based feature (L2). L2 can be calculated as follows for vehicle AV(i) at time epoch (k) which is shown in equation (5)

$$L2_{k=0}^{AV(i)} = \left\| p_{BV}(a_{1,}b_1) - p_{AV}(a_2, b_2) \right\| \quad (5)$$

Where $p_{BV}(a_{1,}b_1)$ and $p_{AV}(a_2, b_2)$represents the position predicting vectors for a Broadcasting vehicle (BV) and the Adjacent Vehicle (AV). $L2_{k=0}^{AV(i)}$is the Euclidian distance measurement between BV and AV.

2. Intersection based Likelihood Level - L3

Vehicle intersection is monitored using the Norbert Bissmeyer method [14]. Intersection Level $L3_k^{AV(i)}$ is the number of times subject vehicle AV(i) intersects other adjacent vehicle's AV(j) communication range in a Moving Window time (MW). The Intersection based Likelihood Level is calculated as shown in equation (6).

$$L3_k^{AV(i)} = \sum_{k=0}^{MW} bin_k^{AV(i,j)} \quad (6)$$

Where $bin_k^{AV(i,j)}$holds the binary result of intersection absence or presence between vehicles (AV(i,j)) at time epoch k.

3.3.3. Behavior based Performance Level (PL) - L4

Performance based features are derived based on the spatial correlation between the Broadcasting vehicle behavior. Behavioral reference could decide the misbehaving act of a vehicle. L4 is calculated as shown in equation (7), where $L4_k^{AV(i)}$ denotes the Performance Level (L4).

$$L4_k^{AV(i)} = \frac{AveragePerformancescore}{ConnectionTimeLength} \quad (7)$$

3.4. Hampel Filter Based Context Reference Building

Context Reference is built using the features derived from the mobility information and Kalman Filter. Four context references are built, namely, Stability Context Reference, Transmission Range Context Reference, Intersection Context Reference, and Performance Context Reference. These context references represent the miscellaneous context. When developing time series models, clear datasets are required. There must not be any lost data or outliers and other anomalies. But VANET datasets have outliers and anomalies due to harsh environments. The Hampel filter is used to build context references without outliers and replace them with more provisional values. For every moving window, the Hampel filter evaluates the median. Windows standard deviation is calculated with the Median Absolute Deviation (MAD) which is shown in figure 2.

$$\sigma = 1.4826\, MAD$$

If any point in the window exceeds more than 3σ apart from the window's median, Hampel filter decides that the point is an outlier and replaces it with a moving window median.

In figure 2, X_s represent the feature data. Hampel filter centers the window to add length. The local median m_i and the standard deviation $\sigma_i$are computed in the recent window of data. The standard deviation of the current window compares with a threshold value α_σ. If the difference between the current feature data and the local median is a greater threshold value of standard deviation, Hampel filter decides that the current feature data X_s as an outlier. Then replaces the outlier with m_i . (i.e) $|X_s - m_i| > \alpha_\sigma \times \sigma_i$.

Let f be a feature extracted from an adjacent vehicle at time epoch k. $x \in L1, L2, L3, L4$. Context reference is denoted as R_k at time epoch k. The context reference is calculated as shown in equation (8).

$$R_k = \begin{cases} m_k = median\ (f) \\ mad_k = 1.4826\ \times median\ \{|f - m_k|\} \\ U_k\ =\ m_k + \ v \times mad_k \\ L_k\ =\ m_k -\ v \times mad_k \end{cases} \quad (8)$$

Where m_k is the median and mad_k is the median absolute deviation. U_k is the upper bound of the Hampel filter and L_k is the lower bound of the Hampel filter. ν represents the threshold. Threshold values are fixed based on the experiments that maximize the accuracy. Context references are constructed and update for every 100 *ms* to capture extremely varying temporal and spatial changes of the context features.

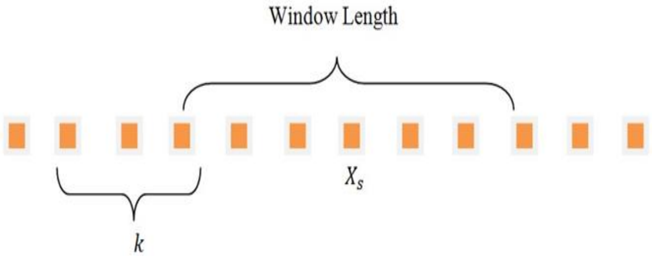


Figure 2 Hampel Filter Based Outlier Detection

3.5. Constructing Miscellaneous VANET Classifiers Based Misbehavior Detection Model (MVC-MDM)

Miscellaneous VANET Classifiers are constructed using the context reference created based on the features extracted from the mobility information and they are used to identify the possible misbehaving vehicles. The first classifier is the stability based classifier and the next two are likelihood based classifiers. Finally, a performance based classifier. Context references are used by the MVC-MDM to distinguish between gentle and malicious vehicles. The fundamental operation of the Classifiers is summarized as follows. The model parameters are taken from the context references, which are used to evaluate the reliability of the mobility information, send by the adjacent vehicles. The standard z- score outlier detection method is used to perform the reliability computation. Standard Z-score is represented as shown in equation (9).

$$z = \frac{x - \mu}{\sigma} \quad (9)$$

Where f is the values of the features, σ is the standard deviation and μ is the arithmetic mean. The standard Z-score method used in this paper is represented as shown in equation (10).

$$z_k^{AV(i)} = \frac{f_{j(k)}^{AV(i)} - m_k^{f_j}}{mad_k^{f_j}} \quad (10)$$

Where $z_k^{AV(i)}$ is the standard Z-score of the adjacent vehicle AV(i). $f_{j(k)}^{v_i}$ is the feature representing the context of a time epoch k. Here μ is replaced with the median$m_k^{f_j}$ and σ is replaced with median absolute deviation $mad_k^{f_j}$. Different statistical classifiers are established considering the context reference and vehicle scores. Varying Miscellaneous VANET Classifiers have been developed based on the vehicle scores and context reference. Each feature is tested with multiple experiments to identify the status of the vehicle behavior. The output of the classifiers is denoted as shown in equation (11).

$$MVC_k^{AV(i)} = \begin{cases} 0 & trusted\ vehicle\ if\ L_k < z_k^{AV(i)} < U_k \\ 1 & otherwise\ misbehaving\ vehicle \end{cases} \quad (11)$$

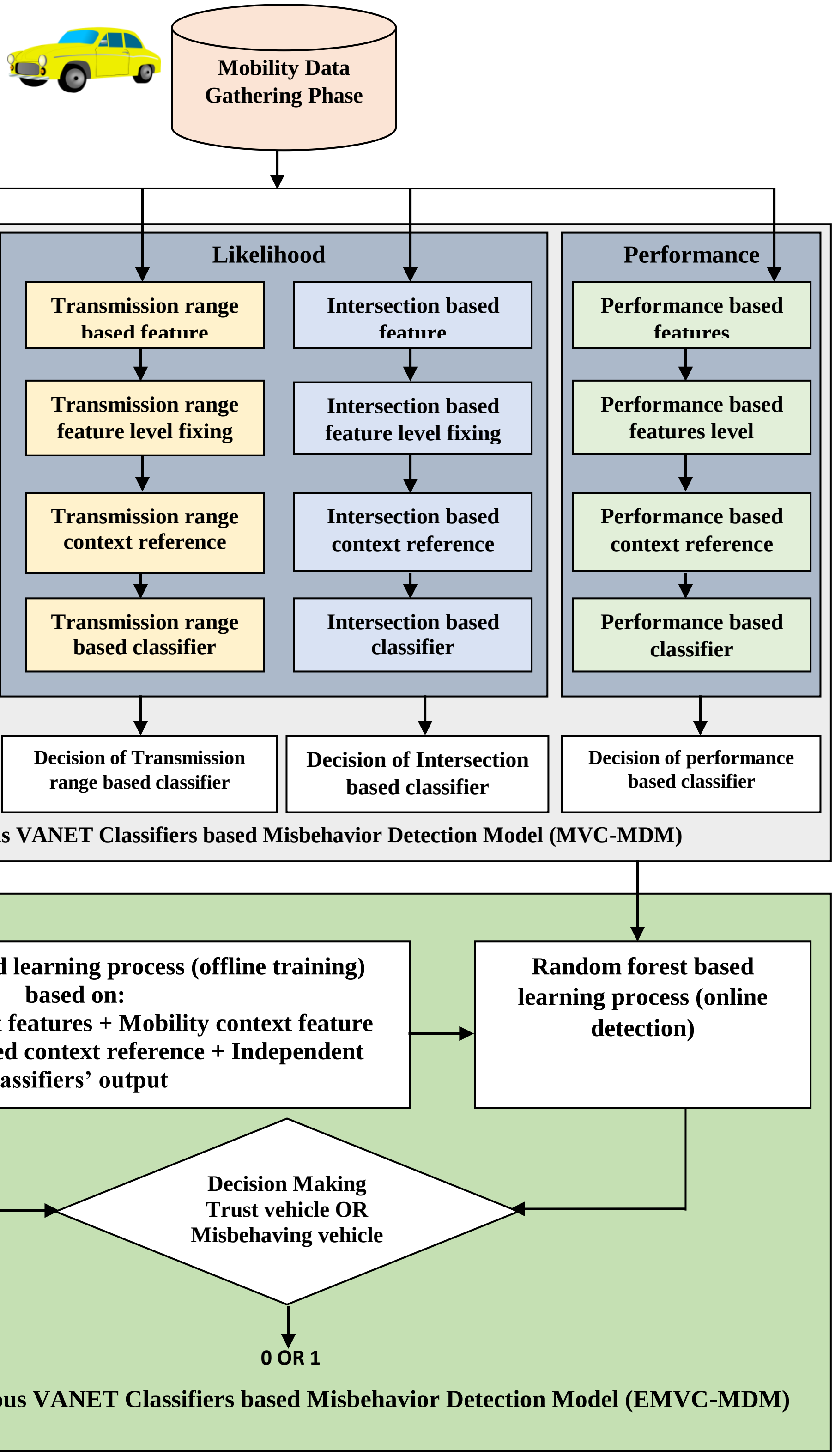

Mobility Data Gathering Phase
Stability
Likelihood
Performance
Stability based features Kalman
Transmission range based feature
Intersection based feature
Performance based features
Stability based feature level fixing
Transmission range feature level fixing
Intersection based feature level fixing
Performance based features level
Stability based context reference
Transmission range context reference
Intersection based context reference
Performance based context reference
Stability based classifier
Transmission range based classifier
Intersection based classifier
Performance based classifier
Decision of Stability based classifier
Decision of Transmission range based classifier
Decision of Intersection based classifier
Decision of performance based classifier
Miscellaneous VANET Classifiers based Misbehavior Detection Model (MVC-MDM)
Random forest based learning process (offline training) based on:
Mobility data context features + Mobility context feature levels + Feature based context reference + Independent classifiers' output
Random forest based learning process (online detection)
Decision Making
Trust vehicle OR
Misbehaving vehicle
0 OR 1
Ensemble Miscellaneous VANET Classifiers based Misbehavior Detection Model (EMVC-MDM)

Figure 3. The proposed MVC-MDM and EMVC-MDM

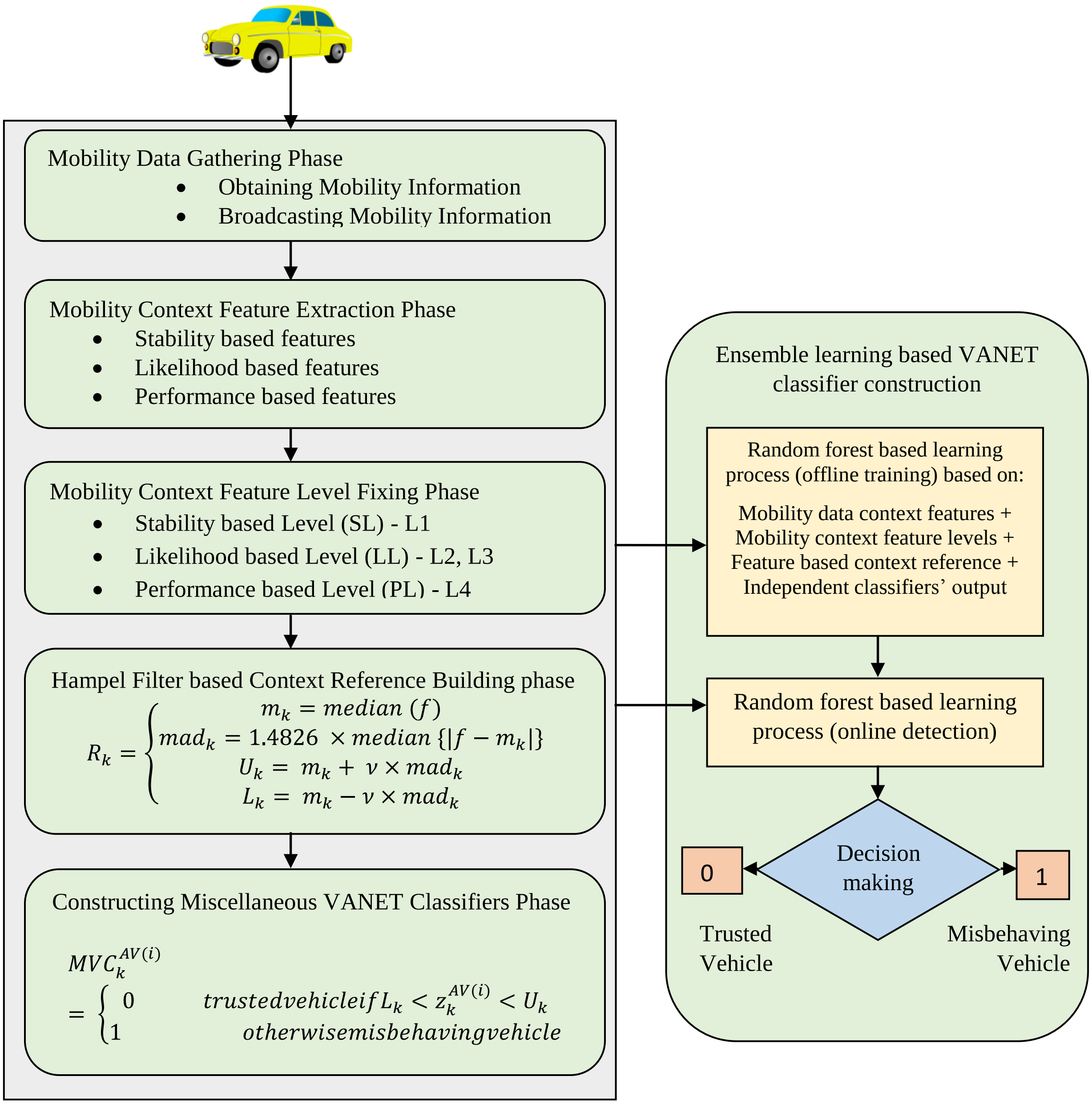

Mobility Data Gathering Phase
Obtaining Mobility Information
Broadcasting Mobility Information
Mobility Context Feature Extraction Phase
Stability based features
Likelihood based features
Performance based features
Mobility Context Feature Level Fixing Phase
Stability based Level (SL) - L1
Likelihood based Level (LL) - L2, L3
Performance based Level (PL) - L4
Hampel Filter based Context Reference Building phase
$R_k = \begin{cases} m_k = median\ (f) \\ mad_k = 1.4826\ \times median\ \{\lvert f - m_k \rvert\} \\ U_k = m_k + v \times mad_k \\ L_k = m_k - v \times mad_k \end{cases}$
Constructing Miscellaneous VANET Classifiers Phase
$MVC_k^{AV(i)} = \begin{cases} 0 & trustedvehicleif\, L_k < z_k^{AV(i)} < U_k \\ 1 & otherwisemisbehavingvehicle \end{cases}$
Ensemble learning based VANET classifier construction
Random forest based learning process (offline training) based on:
Mobility data context features + Mobility context feature levels + Feature based context reference + Independent classifiers' output
Random forest based learning process (online detection)
Decision making
0
1
Trusted Vehicle
Misbehaving Vehicle

Figure 4 Diagrammatic Representation of the Process of the Proposed EMVC-MDM

This procedure is followed to all the context reference model parameters to construct all classifiers such as stability-based classifier, likelihood based classifiers namely transmission range and intersection based classifiers, and performance based classifier. Equation 10 calculates the standard z score of each vehicle according to all the features derived from mobility information. Equation 11 finally evaluates whether the adjacent vehicle AV(i) is a trusted vehicle or misbehaving vehicle. The output is given to logical *OR* operation to identify the final decision. Logical OR derives that if the classifier delivers positive output, then the adjacent vehicle is determined as the misbehaving vehicle. The next section explains the working process of each classifier.

3.6. Ensemble Miscellaneous VANET Classifier based Misbehavior Detection Model (EMVC-MDM)

Pre-detection and vehicle evaluation was done in the previous phase. Miscellaneous VANET Classifiers are built as shown in figure 3. In these classifiers, the input variables are assumed as independent variables. For each mobility feature, a classification rule was developed. Miscellaneous VANET Classifiers do not consider the relationship between the variables. Representing such a classifier design would be vulnerable to context-based attackers. The limitation of independent classifiers is they can only detect the outliers that diverge from the mobility context reference. This results in a successful attack for the attacker who continuously manipulates the mobility data message. To detect such complex attacks, the relationship between the input mobility features is taken into account. For instance, the innovation error, which is the consistency, based feature correlated with the broadcasting rate feature, which is the performance based feature.

An attacker may increase the broadcasting rate and easily spread false information to the adjacent vehicles. If the broadcasting rate were high, the Kalman filter innovation errors would reduce. So the broadcasting rate will not exceed the context reference limits. So to detect sophisticated attacks, the independent classifiers should not ignore the correlation among the mobility features. In VANET, gathering adequate mobility information is a challenging task. Therefore, for identifying previous data patterns, the data received is mined. The relationship between the features remains stationary, so a supervised learning method can be used for predicting complex attacks.

Thus an ensemble classifier learning-based decision making model is created using the random forest method. Random forest decision tree based classifiers are created. Each classifier is trained using random sample features. The majority voting scheme combines the outputs of the classifiers and the misbehaving detection rate is improved using the random forest algorithm. Figure 3 shows the working procedure of the proposed ensemble learning model based on the Miscellaneous VANET Classifiers.

There are two main tasks done in the EMVC - MDM model, Offline training and online prediction operation. The dataset is prepared first offline training by injecting dynamic noises to simulate the VANET noise environment. Under different communication scenarios, the prepared datasets are trained. The next step is to inject complex attack types into the vehicle trajectories, which creates faulty mobility information. The features are derived based on the context and the context references are built using the features derived from the context. The mobility context feature levels and the outputs from the classifiers are used as the input to the random forest algorithm. By combining the random forest based classifiers and the Miscellaneous VANET Classifiers model the online operation for prediction is carried out. The outputs of the final RF based ensemble classifiers were analyzed under the average weight for decision making. The complete flow of the proposed EMVC-MDM model is shown in figure 4.

3.7 Misbehavior decision phase

The outputs of the RF based ensemble classifiers decide the final decision using the following equation (12).

$$M_k^{AV(i)} = \begin{cases} 0 & trusted\ vehicle\ if\ \frac{\Sigma\, c_w * c_o}{\Sigma\, c_w} > C \\ 1 & misbehaving\ vehicle \end{cases} \quad (12)$$

Where $\mathrm{M}_\mathrm{k}^{\mathrm{AV(i)}}$is the final decision. $\mathrm{c_w}$ is the weight and $\mathrm{c_o}$ is the output of the classifier j. 'C' is the weight of the classifier.

4. PERFORMANCE EVALUATION

To evaluate the performance of the proposed MVC and EMVC method, an extensive simulation has been conducted. First, all the common steps in experimenting with the misbehavior detection model are followed as communication simulation, creating misbehavior environment, dynamic noise injection, selecting an appropriate dataset, and processing the datasets. For simulation purposes, *Network Simulator (NS2)* is used to create the VANET environment, and Simulator for Urban Mobility environment *(SUMO)* is used as the traffic simulator. *MATLAB* is used for computation purposes. *MATLAB* simulates the communication channel and dynamic noises to predict misbehaving vehicles.

2 - Ray ground reflection model is used to create VANET wireless channel. SUMO- Simulation for Urban Mobility tool, is used for traffic simulation. The *osm* file of the sample city view is created from the Open Street Map tool. The communication range between each vehicle is about 1 km. the maximum speed of the vehicle is 20 *km/s.* The simulation area is fixed as 1000 × 1000 *meters*. 200 vehicles are used for the simulation. Channel bandwidth is about 3 *Mbps*. Table 2 shows the simulation parameters values used in this paper for simulation.

4.1. Preprocessing of Datasets

In this paper, datasets are gathered from Next Generation Simulation (NGSIM), which is a dataset of the real world traffic situation. NGSIM provides the real world vehicles' trajectories such as position, velocity, acceleration, heading direction, etc. These datasets are recorded at each *100 ms* by the vehicle sensors and positioning devices [21]. The dataset is subdivided into four clusters that include all types of driver activities. Each vehicle's temporal features, spatial features, and behavioral features were selected. These features are used to find their variance and covariance. Clusters are formed depending upon the driver's behavior such as car flowing, random flowing, lane changing, and free flowing [22].

Simulation Parameters	**Values**
Max Speed	20 m/s
Communication Range	1km
Channel Bandwidth	3Mbps

Simulation time	500 seconds
Simulation Area	1000×1000 meters
Number of Vehicles	200
Wireless Channel	2-ray channel
Traffic Simulator	SUMO
Network Simulator	NS 2.34
Map Model	OSM
Wireless Channel	2-ray channel

Table 2 Simulation Settings

4.2. Injecting Environmental Noises

In a VANET environment, vehicles subject to different noises. Noise is the error or an undesired random distribution of information. In this paper, three different noise types are used such as stationary white noise, dynamic non-stationary noise, and correlated noise based on the environmental conditions. *Noise1* is considered as the combination of stationary white noise with zero means; *Noise2* is considered as the non-stationary white noise with time-varying variance, and *Noise3* is a correlated noise. Using different noise types helps in proving the effectiveness of the proposed algorithm. Stationary White noise occurs when vehicles travel on highways and rural environments [23]. In contrast, non-stationary white noise occurs during traveling in harsh environments such as urban areas, cloudy or misty weathers, and tunnels. In those situations, signals are affected and damaged. Many researchers in the VANET positioning context analyzed these noise types. *Noise1* generally occurs under Line of Sight (LoS) measurement conditions such as Global Navigation Satellite System (GNSS) based positioning in a blue sky environment [24]. *Noise1* appears on the highway or rural environments. *Noise2* occurs when vehicles travel in harsh environments such as clouds and trees. Cloud water bodies could absorb the signal. *Noise3* happens explicitly, such as in tunnels, under bridges, and in the middle of town areas. The moving window size is *120*, and the sampling rate is *100 milliseconds* [25].

4.3. Evaluating Message Loss

VANET applications nearly depend on the effect of the message loss ratio. To ensure that, the simulation for the proposed method is tested with ten communication scenarios, which have various traffic densities and produce a different message loss ratio [26]. Sixteen various traffic datasets have been conducted for each communication scenario. Entirely *160* experiments were carried out to evaluate the proposed misbehavior detection model. *NS2* is used as the network simulator, which connects the vehicles using WAVE communication. The maximum communication range between two adjacent vehicles is *1km* and the broadcasting frequency is *10Hz.* At the starting of Control Channel Interval (*CCHI*), each vehicle broadcasts a new beacon with a mobility message [27]. To evaluate the message loss ratio, the delay of the message is considered as a random variable using Poisson distribution.

4.4. Creating Misbehaving Environment

For assessing the misbehavior detection model in VANET, creating a misbehavior environment is the common step to be followed. Two types of misbehaving environments can be created such as attackers and flawed vehicles [28]. In this paper, two types of attacks are used to create the misbehavior environment such as common attacks and complicated attacks. Common attacks alter the mobility information, but not altering the context of the mobility information. An example of a common attack is fake position and positioning noises [29]. Complicated attacks are keen on mobility context. An example of a complicated attack is creating fake events, sudden brake applying, and prompt the vehicle to take a diversion. Flawed vehicles are of two types, hardware flaws, and software flaws. Vehicle flaw attacks are ignored in this paper.

4.5. Performance Metrics

In this paper, five performance metrics are used to compare the proposed method and the existing methods such as detection rate, detection accuracy, F-measure, specificity, and precision. The existing misbehavior detection models to compare with others generally use these metrics. Misbehavior Detection rate is the ratio of exactly detected attackers by a total number of actual attackers [30]. The detection rate is also called sensitivity or Recall. Misbehavior Detection accuracy is the ratio of the total number of exactly detected vehicles with a total number of vehicles. Otherwise, detection accuracy is also calculated as the sum of true positive detection and false positive detection divided by the total true positive, false positive, true negative, and false negative detections [31]. Precision refers to the rate of a total number of exactly detected attackers by the total number of vehicles classified as attackers. Specificity is the ratio between how many vehicles were exactly detected as misbehaving vehicles on how many vehicles are misbehaving. F-measure is to combine the two scores of recall and precision into a single score [32]. The false-positive rate is the probability of a false rate, when the true value is negative, the positive result will be given. These performance metrics are calculated using the following equations (13) to (18).

$$Detection\ Rate\ (Recall) = \frac{TP}{TP + FN} \tag{13}$$

$$Detection\ Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$Specificity = \frac{TN}{TN + FP} \tag{16}$$

$$F\ measure = \frac{2\ \times precision\ \times recall}{precision + recall} \tag{17}$$

$$False\ Positive\ Rate = \frac{FP}{FP + TN} \tag{18}$$

Where TP is the True Positive Rate, TN is True Negative Rate, FP False Positive Rate, FN False Negative Rate [30].

5. RESULTS AND DISCUSSION

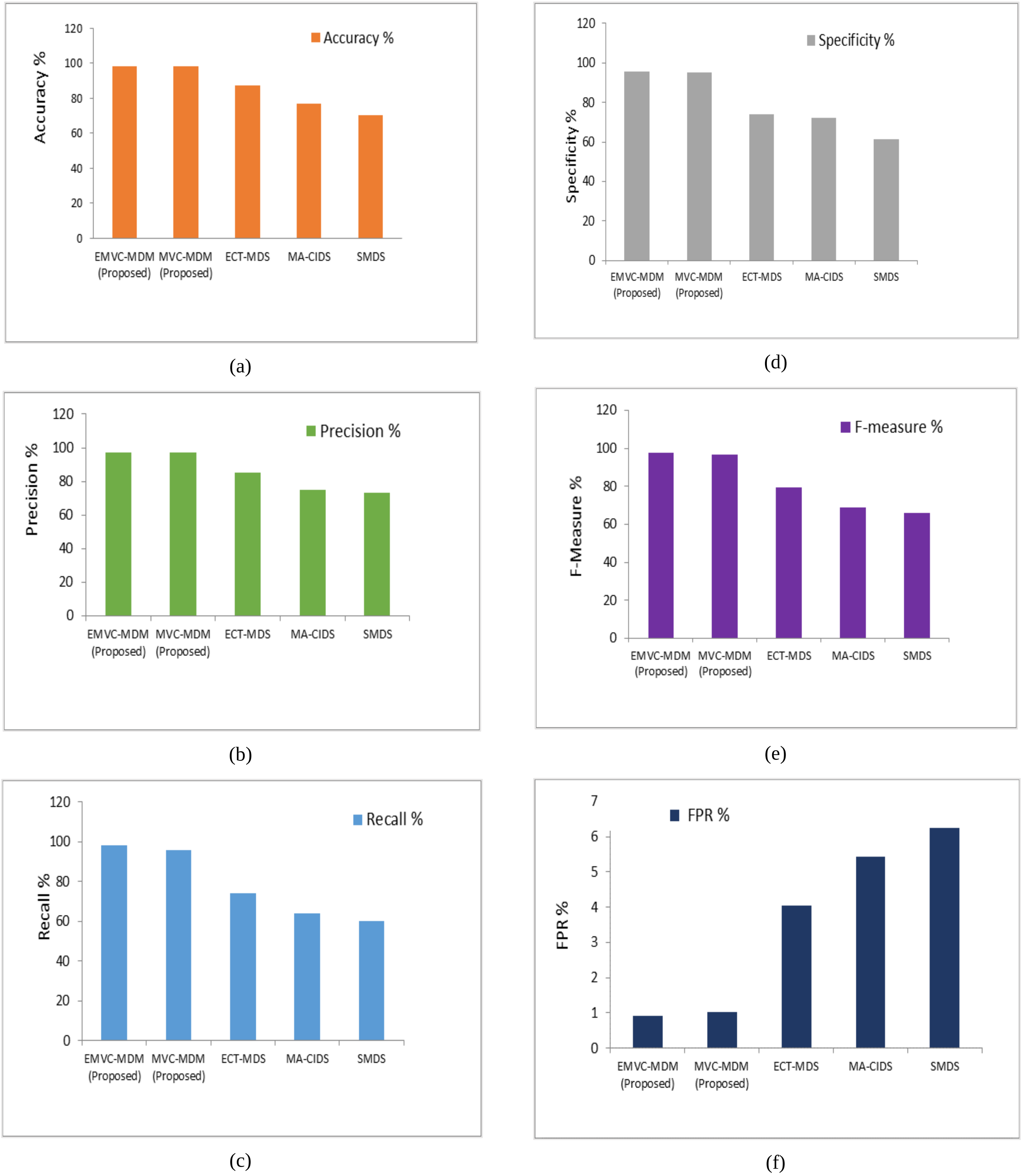
This section briefly discusses the performance of the proposed MDS compared with the existing MDS. The proposed MDS, Miscellaneous VANET Classifiers based Misbehavior Detection Model (MVC-MDM) and Ensemble Miscellaneous Classifiers based Misbehavior Detection Model (EMVC-MDM) are weighed against the existing MDS such as ECT-MDS [16], MA-CIDS [17], Stubing's MDS (SMDS) [18] in terms of various performance metrics. SMDS is used as the baseline for examining the performance of the proposed MDS. This section is divided into two parts, where, first part explains the effectiveness of the proposed misbehavior detection model and the existing models. The second part explains the reliability achieved by the proposed and the existing misbehavior detection model.

5.1. Results of Effectiveness

Table 3 shows the results obtained by the proposed and existing MDS. This table shows that the proposed EMVC-MDM has attained the best accuracy results in terms of misbehavior detection accuracy. Figure 5 shows the diagrammatic representation of the Effectiveness of the proposed MVC-MDM and EMVC-MDM in detail. The x-axis holds the proposed and the existing models and the y-axis holds the ratio obtained from that misbehavior detection models. EMVC-MDM achieves about 98.34% of accuracy, which means that 98.34% of vehicles exactly identified as misbehaving vehicles. Other proposed independent classifiers MVC-MDM could achieve 98.02% of misbehavior detection accuracy. The existing methods such as ECT-MDS, MA-MDS, and SMDS obtains 87.21 %, 76.76%, and 70.21% ratio of accuracy respectively which is shown in figure 5.a. In Table 3, the precision ratio attained by the proposed EMVC-MDS is 97.21% and proposed independent classifiers MVC-MDM attain 96.02% of precision. Existing MDS such as ECT-MDS brings 85.30 precision, MA-CIDS brings 74.72% precision and SMDS brings 73.15% precision. These results show that the proposed EMVC-MDM achieves a high precision ratio shown in figure 5.b. The proposed EMVC-MDM results highest recall ratio of 98.24% and MVC-MDM brings a 96.02% Recall value. But the existing method results in lower Recall values such as 74.14% by ECT-MDS, 63.86% by MA-CIDS, and 60.21% by SMDS. In figure 5.c, the results show, the proposed MDS achieves the highest Recall value among all the existing MDS. Specificity is one of the important performance metrics, which shows the effectiveness of the proposed model. Such that 95.64% of specificity is gained by the proposed EMVC-MDM and MVC-MDM gains 95.21 % of specificity which is the highest one among the existing MDS. ECT-MDS achieves 74.01%, MA-CIDS achieve 72.15% and SMDS achieves 61.38% specificity which is shown in figure 5.d. The result of the F-measure tells how well the proposed model works on the misbehavior detecting process. Such that the proposed EMVC-MDM reaches 97.72% of F-measure. Another proposed model, MVC-MDM scores 96.46% of F-measure. However, the existing method could not reach such an F-measure value. ECT-MDS scores 79.32%, MA-CIDS scores 68.86% and SMDS scores 66.05% of F-measure value, which is shown in figure 5.e. An effective misbehavior detection model should minimize the False Positive Rate (FPR) because the false positive rate will degrade the performance of the misbehavior detection method. Based on this fact, the proposed EMVC-MDM minimizes the FPR up to 0.92% and MVC-MDM minimizes the FPR up to 1.01%. Nevertheless, the existing methods show 4.04% FPR by ECT-MDS, 5.43% by MA-CIDS, and 6.26% by SMDS, which is shown in, figure 5.f.

Model	Accuracy %	Precision %	Recall %	Specificity %	F-measure %	FPR %
EMVC-MDM (Proposed)	98.34	97.21	98.24	95.64	97.72	0.92
MVC-MDM (Proposed)	98.02	96.92	96.02	95.21	96.46	1.01
ECT-MDS	87.21	85.30	74.14	74.01	79.32	4.04
MA-CIDS	76.76	74.72	63.86	72.15	68.86	5.43
SMDS	70.21	73.15	60.21	61.38	66.05	6.26

Table 3 Results of Effectiveness

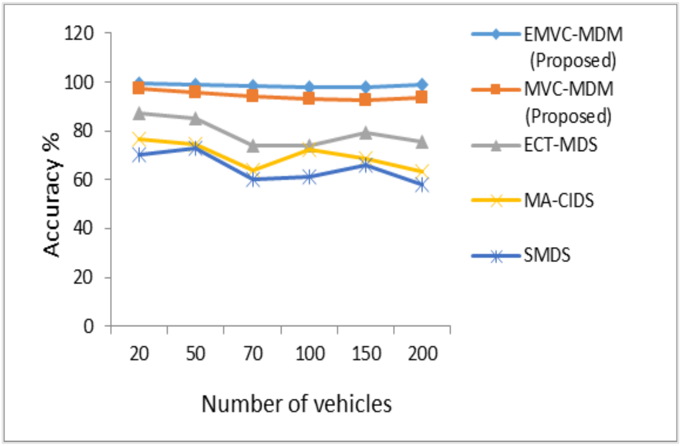


(a) (d) (b) (e) (c) (f)

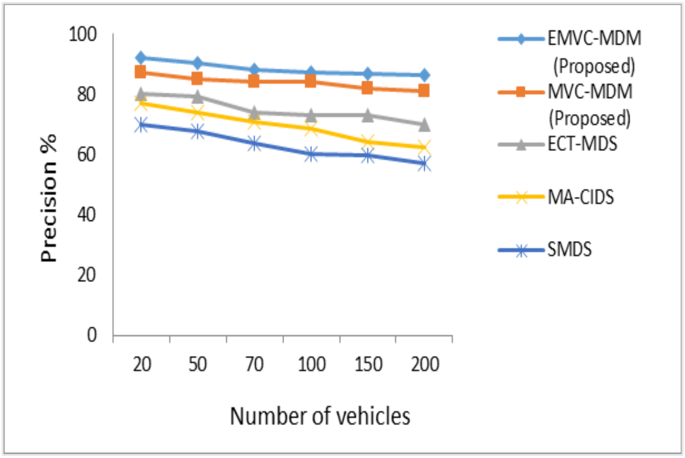
Figure 5 Results of effectiveness in terms of (a) Accuracy, (b) Precision, (c) Recall, (d) Specificity, (e) F-measure and, (f)False Positive Rate (FPR)

5.2. Results of Reliability

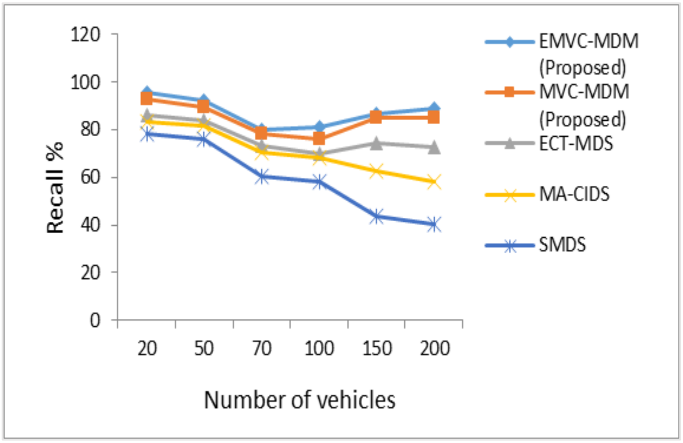
The reliability is the robustness of the misbehavior detection models in various vehicle density scenarios. In this paper, 200 VANET nodes (vehicles) have been used. The obtained results are shown in Table 4. Figure 6 shows the reliability obtained by the misbehavior detection models at various vehicle density scenarios. In these figures, the x-axis holds the number of vehicles to represent the various vehicle densities. Y-axis holds the ratio of the performance metrics achieved by the MDS. In figure 6.a, at different vehicle densities such as 20, 50, 70, 100, 150, and 200, the proposed EMVC-MDM sets the average accuracy value of 98.50%. The proposed EMVC-MDM maintains the highest accuracy level compared to the other MDS. The proposed MVC-MDM evenly maintains the average accuracy level of 94.45%. However, the existing MDS such as ECT-MDS, MA-CIDS, and SMDS, a lower accuracy level of 79.22%, 69.93%, and 64.87%. Figure 6.b shows the precision value achieved by the MDS at different vehicle densities. The proposed EMVC-MDM attains an average of 88.56% of precision ratio. MVC-MDM achieves an average of 84.06% precision ratio at various densities such as 20, 50, 70, 100,150, and 200 vehicles. Meanwhile, ECT-MDS, MA-CIDS, and SMDS attain an average of 74.95%, 69.57%, and 63.17% of precision ratio at various vehicle densities. In figure 6.c, it shows the Recall ratio achieved by the MDS at various vehicle densities. The proposed EMVC-MDM attains 87.59% of Recall ratio and MVC-MDM attains 84.46% of Recall ratio. Whereas the existing MDS such as ECT-MDS attain an average of 76.75%, MA-CIDS attain 70.80% and SMDS attains 59.43% of the Recall ratio. All the three existing MDS Recall ratio is lower than the proposed MDS. This shows the effectiveness of the proposed MDS. The specificity ratio is shown in figure 6.d at various vehicle densities. The specificity values obtained by ECT-MDS, MA-CIDS, and SMDS are 72.42%, 64.54%, and 59.13%, respectively, whereas the proposed EMVC-MDM and MVC-MDM achieve higher specificity values such as 85.35% and 74.87% respectively. Figure 6.e shows the F-measure ratio obtained from the MDS. EMVC-MDM obtains an average of 83.27% F-measure ratio and MVC-MDM obtains an average of 79.18 F-measure at various vehicle densities. The existing methods such as ECT-MDS gains average of 74.09% F-measure, MA-CIDS gains an average of 68.69% of F-measure and SMDS gains an average of 55% of F-measure ratio of various vehicle densities. In figure 6.f, the False Positive Rate (FPR) produced by each MDS is shown. In this figure, it is seen that the proposed EMVC-MDM produces an average of 1.30% of FPR and MVC-MDM produces 2.22% of FPR, which is the lowest FPR among all the existing MDS. ECT-MDS produces an average of 2.64% of FPR, MA-CIDS produces an average of 4.01% of FPR and SMDS produces an average of 5.22% of FPR at various vehicle densities. Table 4 shows the results of reliability obtained by the MDS. From these results, it is proved that the proposed EMVC-MDM and MVC-MDM models perform better than the existing misbehavior detection models such as ECT-MDS, MA-CIDS, and SMDS.



(a)

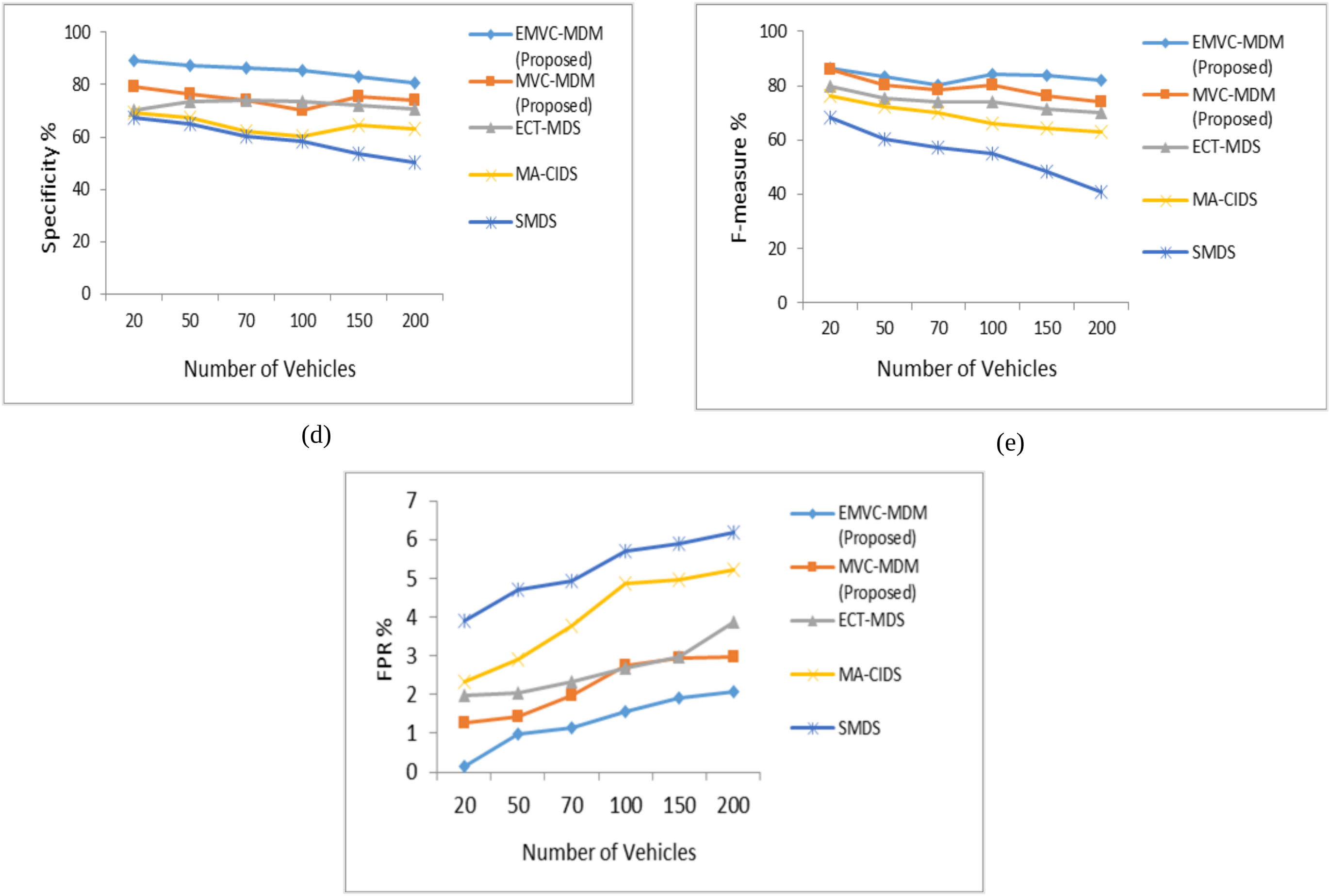


(b)



(C)

RESEARCH ARTICLE



(d)

(e)

(f)

Figure 6 Results of Reliability in terms of (a) Accuracy (b) Precision (c) Recall (d) Specificity (e) F-measure and (f) False Positive Rate (FPR)

Model	Accuracy %	Precision %	Recall %	Specificity %	F-measure %	FPR %
EMVC-MDM (Proposed)	98.5	88.56	87.59	85.35	83.27	1.3
MVC-MDM (Proposed)	94.45	84.06	84.46	74.87	79.18	2.22
ECT-MDS	79.22	74.955	76.75	72.42	74.09	2.64
MA-CIDS	69.93	69.575	70.8	64.54	68.69	4.01
SMDS	64.87	63.17	59.43	59.13	55.06	5.22

Table 4 Results of Reliability

6. CONCLUSION

In this paper, Ensemble Miscellaneous VANET Classifiers based Misbehavior Detection Model (EMVC-MDM) is proposed along with the independent classifiers namely, Miscellaneous VANET Classifiers based Misbehavior Detection Model (MVC-MDM). The main concept of the proposed MDS is to classify the vehicles, according to the context reference to identify the misbehaving vehicle. The ensemble method integrates the independent classifiers using the Random forest ensemble method. Classifiers are created using the context reference built based on the level of context features extracted from the mobility information received by the adjacent vehicle. Both supervised and unsupervised methods were used to build the VANET classifiers. Hampel filter based outlier detection method along with standard Z-score method was used to construct the context reference. The Random forest ensemble method combines the correlations among the classifiers and improves their efficiency in the misbehavior detection process. The proposed MDS shows a significant enhancement in terms of misbehavior detection compared to the existing misbehavior detection systems. The EMVC-MDM and MVC-MDM show effective adaptability and reliability under unstable communication and dynamic noise environments. The overall performance of the proposed. EMVC-MDM and MVC-MDM are enhanced by an average of more than 15% better results than the existing MDS.

As the future process, the supervised learning methods can be replaced by deep learning methods and Artificial Neural Network (ANN) concepts. To improve the data transfer speed. 5G technology can be used in the best simulation environments. The proposed EMVC-MDM relies on detecting short-term misbehaviors (only the mobility information shared by the adjacent vehicles.). In the future, the proposed method can be enhanced to detect long-term misbehaviors beyond mobility information.

REFERENCES

[1] Lim, Kiho, Kastuv M. Tuladhar, and Hyunbum Kim. Detecting location spoofing using ADAS sensors in VANETs, in 2019 16th IEEE annual consumer communications & networking conference (CCNC), IEEE, 2019, pp. 1-4.

[2] H. Vahdat-Nejad, A. Ramazani, T. Mohammadi, and W. Mansoor, A survey on context-aware vehicular network applications, Vehicular Communication, vol. 3, Jan. 2016, pp. 43-57.

[3] O. A. Wahab, A. Mourad, H. Otrok, and J. Bentahar, CEAP: SVM based intelligent detection model for clustered vehicular ad hoc networks, Expert System Applications, vol. 50, May 2016, pp. 40-54.

[4] R.W. van der Heijden, S. Dietzel, T. Leinmüller, and F. Kargl, Survey on misbehavior detection in cooperative intelligent transportation systems, IEEE Commun. Surveys Tuts., vol. 21, no. 1, 4th Quart, 2018, pp. 779-811.

[5] F. Sakiz and S. Sen, A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV, Ad Hoc Network., vol. 61, Jun. 2017, pp. 33-50.

[6] M. N. Mejri, J. Ben-Othman, and M. Hamdi, Survey on VANET security challenges and possible cryptographic solutions, Vehicular Communication, vol. 1, no. 2, Apr. 2014, pp. 53-66.

[7] S. Sumithra and R. Vadivel, An overview of various trust models for VANET security establishment, 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, 2018, pp. 1-7.

[8] J. Wang, Y. Shao, Y. Ge, and R. Yu, A survey of vehicle to everything (V2X) testing, Sensors, vol. 19, Jan. 2019, pp. 334.

[9] R. K. Pearson, Y. Neuvo, J. Astola, and M. Gabbouj, Generalized Hampel filters, EURASIP J. Adv. Signal Process., vol. 2016, pp. 87.

[10] U. Khan, S. Agrawal, and S. Silakari, A detailed survey on misbehavior node detection techniques in vehicular ad Hoc networks, in Information Systems Design and Intelligent Applications (Advances in Intelligent Systems and Computing), New Delhi, India: Springer, vol. 339, 2015, pp. 11-19.

[11] F. A. Ghaleb, A. Zainal, A. M. Rassam, and F. Saeed, ``Driving-situation aware adaptive broadcasting rate scheme for vehicular adhoc network, Journal of Intelligent Fuzzy Systems, vol. 35, 2018, pp. 423-438.

[12] X. Y. Tian, Y. H. Liu, J. Wang, W. W. Deng, and H. Oh, Computational security for context-awareness in vehicular ad-hoc networks, IEEE Access, vol. 4, 2016, pp. 5268-5279.

[13] S. Dietzel, J. Petit, G. Heijenk, and F. Kargl, Graph-based metrics for insider attack detection in VANET multihop data dissemination protocols, IEEE Transaction on Vehicular Technology, vol. 62, no. 4, May 2013, pp. 1505-1518.

[14] S. Ruj, M. A. Cavenaghi, Z. Huang, A. Nayak, and I. Stojmenovic, On data-centric misbehavior detection in VANETs, in Proc. IEEE Vehicular Technology. Conf. (VTC Fall), Sep. 2011, pp. 1-5.

[15] S. Sumithra and R. Vadivel, NB-FTBM model for entity trust evaluation in vehicular ad hoc network security, 2nd International Conference on Ubiquitous Communications and Network Computing, Springer, Cham, 2019, pp. 173-187.

[16] N. Bissmeyer, W. Michael, and K. Frank, Misbehavior detection and attacker identification in vehicular ad-hoc networks, Tech. Univ. Darmstadt, Darmstadt, Germany, Tech. Rep., 2014.

[17] A Ghaleb, Fuad, Faisal Saeed, Mohammad Al-Sarem, Bander Ali Saleh Al-rimy, Wadii Boulila, A. E. M. Eljialy, Khalid Aloufi, and Ma1111moun Alazab. Misbehavior-aware on-demand collaborative intrusion detection system using distributed ensemble learning for VANET, Electronics, vol.9, no. 9, 2020, pp.1411.

[18] H. Stubing, Car-to-X communication: System architecture and applications, in Multilayered Security and Privacy Protection in Car-to-XNetworks. Wiesbaden, Germany: Springer, 2013, pp. 9-19.

[19] Zhang, Chunhua, Kangqiang Chen, Xin Zeng, and Xiaoping Xue. "Misbehavior detection based on support vector machine and Dempster-Shafer theory of evidence in VANETs." IEEE Access, vol.6, 2018: 59860-59870.

[20] S. Sumithra and R. Vadivel, Optimal Innovation-Based Adaptive Estimation Kalman Filter For Measuring Noise Uncertainty During Vehicle Positioning In VANET, International Journal of Applied Mathematics and Computer Science (AMCS), Vol. 31, No. 1, March 2021.

[21] Kamel, Joseph, Mohammad Raashid Ansari, Jonathan Petit, Arnaud Kaiser, Ines Ben Jemaa, and Pascal Urien. "Simulation framework for misbehavior detection in vehicular networks." IEEE transactions on vehicular technology, vol.69, no. 6, 2020, pp.6631-6643.

[22] K. Zaidi, M. B. Milojevic, V. Rakocevic, A. Nallanathan, and M. Rajarajan, Host-based intrusion detection for VANETs: A statistical approach to rogue node detection, IEEE Transaction on Vehicular Technology, vol. 65, no. 8, Aug. 2016, pp. 6703-6714.

[23] Ghaleb, Fuad A., Anazida Zainal, Murad A. Rassam, and Fathey Mohammed, An effective misbehavior detection model using artificial neural network for vehicular ad hoc network applications, In 2017 IEEE Conference on Application, Information and Network Security (AINS), IEEE, 2017, pp. 13-18.

[24] S. A. Soleymani, A. H. Abdullah,W. H. Hassan, M. H. Anisi, S. Goudarzi, and M. A. R. Baee, Trust management in vehicular ad hoc network: A systematic review, EURASIP J. Wireless Communication Network, vol. 1, Dec. 2015, pp. 146.

[25] Y. Zhang, L. Lazos, and W. Kozma, AMD: Audit-based misbehavior detection in wireless ad hoc networks, IEEE Transaction on Mobile Computing., vol. 15, no. 8, Aug. 2016, pp. 1893-1907.

[26] Dietzel, Stefan, Rens van der Heijden, Hendrik Decke, and Frank Kargl, A flexible, subjective logic-based framework for misbehavior detection in V2V networks, In Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, IEEE, 2014, pp. 1-6.

[27] N. Lyamin, A. Vinel, M. Jonsson, and B. Bellalta, ``Cooperative awareness in VANETs: On ETSI EN 302 637-2 performance," IEEE Transaction on Vehicular Technology, vol. 67, no. 1, Jan. 2018, pp. 17-28.

[28] Sakiz, F.; Sen, S, A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV, Ad Hoc Networks, vol.61, 2017, pp. 33–50.

[29] Van der Heijden, R.W.; Stefan, D.; Tim, L.; Frank, K, Survey on misbehavior detection in cooperative intelligent transportation systems, IEEE Communications Surveys & Tutorials, vol.21, no.1, 2019, pp.779–811.

[30] Ho, Yao-Hua, Chun-Han Lin, and Ling-Jyh Chen, On-demand misbehavior detection for vehicular ad hoc network, International Journal of Distributed Sensor Networks, vol.12, no. 10, 2016, 1550147716673928.

[31] Ghaleb, Fuad A., Mohd Aizaini Maarof, Anazida Zainal, Murad A. Rassam, Faisal Saeed, and Mohammed Alsaedi, Context-aware data-centric misbehavior detection scheme for vehicular ad hoc networks using sequential analysis of the temporal and spatial correlation of the consistency between the cooperative awareness messages, Vehicular Communications, vol.20, 2019, 100186.

[32] Erskine, Samuel Kofi, and Khaled M. Elleithy, Real-time detection of DoS attacks in IEEE 802.11 p using fog computing for a secure intelligent vehicular network, Electronics 8, no. 7, 2019, pp.776.

Authors

S.Sumithra is a Ph.D. candidate at Bharathiar University, Tamil Nadu, India. She started her research in 2017 in the area of computer science. Her current research interest includes the Intelligent Transport System, Deep learning, Artificial Intelligence, and Vehicular Ad Hoc Network. She received her M.Sc Degree in Information Technology and M.Phil degree in Computer Science from Bharathiar University in 2015 and 2017 respectively. She has started her research in advanced networking. She is also interested in 5G wireless solutions and optical networking.

Dr. R. Vadivel is an Assistant Professor in the Department of Information Technology, Bharathiar University, Tamil Nadu, India. He received his Ph.D. degree in Computer Science from Manonmaniam Sundaranar University in the year 2013. He obtained his Diploma in Electronics and Communication Engineering from State Board of Technical Education in the year 1999, B.E., Degree in Computer Science and Engineering from Periyar University in the year 2002, M.E., a degree in Computer Science and Engineering from Annamalai University in the year 2007. He had published over 40 journal papers and over 30 conference papers both at the National and International levels. His areas of interest include Computer Networks, Network Security, Information Security, etc.