




Article

Ensemble-Based Hybrid Context-Aware Misbehavior Detection Model for Vehicular Ad Hoc Network

Fuad A. Ghaleb ^{1,*}, Mohd Aizaini Maarof ^{1,*}, Anazida Zainal ¹, Bander Ali Saleh Al-rimy ¹, Abdullah Alsaedi ² and Wadii Boulila ^{2,3}

¹ Cyber Threat Intelligence Lab (CTIL), School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, Johor 81310, Malaysia; anazida@utm.my (A.Z.); bnder321@gmail.com (B.A.S.A.)

² College of Computer Science and Engineering, Taibah University, Medina 42353, Saudi Arabia; aasaedi@taibahu.edu.sa (A.A.); wadii.boulila@riadi.rnu.tn (W.B.)

³ RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba 2010, Tunisia

* Correspondence: aafuad@utm.my (F.A.G.); aizaini@utm.my (M.A.M.);
Tel.: +60-11-273-46783 (F.A.G.); +60-19-700-0705 (M.A.M.)

Received: 3 November 2019; Accepted: 29 November 2019; Published: 1 December 2019



Abstract: Life-saving decisions in vehicular ad hoc networks (VANETs) depend on the availability of highly accurate, up-to-date, and reliable data exchanged by neighboring vehicles. However, spreading inaccurate, unreliable, and false data by intruders create traffic illusions that may cause loss of lives and assets. Although several solutions for misbehavior detection have been proposed to address these issues, those solutions lack adequate representation and the adaptability to vehicular context. The use of predefined static thresholds and lack of comprehensive context representation have rendered the existing solutions limited to specific scenarios and attack types, which impedes their generalizability. This paper addresses these limitations by proposing an ensemble-based hybrid context-aware misbehavior detection system (EHCA-MDS) model. EHCA-MDS has been developed in four phases, as follows. The static thresholds have been replaced by dynamic ones created on the fly by analyzing the spatial and temporal properties of the mobility information collected from neighboring vehicles. Kalman filter-based algorithms were used to collect the mobility information of neighboring vehicles. Three sets of features were then derived, each of which has a different perspective, namely data consistency, data plausibility, and vehicle behavior. These features were used to construct a dynamic context reference using the Hampel filter. The Hampel-based z-score was used to evaluate the vehicles based on their behavioral activities, data consistency, and plausibility. For comprehensive features representation, multifaceted, non-parametric-based statistical classifiers were constructed and updated online using a Hampel filter-based algorithm. For accurate representation, the output of the statistical classifiers, vehicles' scores, context reference parameters, and the derived features were used as input to an ensemble learning-based algorithm. Such representation helps to identify the misbehaving vehicles more effectively. The proposed EHCA-MDS model was evaluated in the presence of different types of misbehaving vehicles under different context scenarios through extensive simulations, utilizing a real-world traffic dataset. The results show that the accuracy and robustness of the proposed EHCA-MDS under different vehicular dynamic context scenarios were higher than existing solutions, which confirms its feasibility and effectiveness to improve the performance of VANET critical applications.

Keywords: context-aware misbehavior detection; vehicular ad hoc network; VANET; hybrid; false information; illusion attack; Hampel filter; Kalman filter; consistency; plausibility

1. Introduction

Road collisions are increasing, and they are being expected to be the fifth leading cause of death by 2030 [1,2]. Annually, millions of people lose their lives on roads worldwide due to traffic accidents [1], with 40 times more suffering from injuries. These accidents are also the main cause of traffic congestion, which in turn has a great impact on the economy [3,4], and billions of dollars are lost due to the treatment of injuries, loss of property, lost working hours, and high fuel consumption [5]. Several studies reveal that more than 95% of accidents are attributed to human error that can be avoided if the drivers had been warned half a second beforehand [6]. Vehicular ad hoc networks (VANETs) have emerged, mainly to reduce those accidents, improve road safety and traffic efficiency, and provide user comfort [7]. Vehicles with hundreds of sensors and communication capabilities share their real-time information about their movement state, road conditions, and traffic situation [8,9]. Mobility information is the main building block of VANET applications and services to create hazard-free roads, which optimize the traffic efficiency, and improve network agility. Vehicles use wireless access in vehicular environments (WAVE) devices, also called on-board units (OBU), to communicate with each other in an ad-hoc manner and/or with Road Sites Units (RSUs) in the infrastructure. These devices utilize the dedicated short range communication (DSRC) protocol to form vehicle-to-vehicle (V2V), and/or vehicle-to-infrastructure (V2I) links.

By utilizing VANET's capabilities, a wide range of applications can be provided for safety, traffic efficiency, and commercial applications [10]. Many applications have received attention from research and industry, such as cooperative collision warning (CCWS) [11], intersection safety [12], cooperative driving [13], cooperative adaptive cross-control (CACC) [14], and driver assistance systems (ADAS) [15]. These applications have a huge potential to save tens of thousands of lives and billions of dollars each year [16]. An extensive review of VANET applications can be found in References [6,17]. With such applications, vehicle automation will become a reality, such that human error will be minimized, and the vehicles can navigate safely even under hazardous and dangerous traffic conditions such as fog, black ice, and accidents. The performance of most VANET applications depends on the availability of continuous, accurate, and up-to-date vehicle mobility information, including position information, speed information, and direction information [18]. In VANET, vehicles cooperatively share their mobility information to facilitate the operation of applications. However, VANET is vulnerable to many types of attacks that can disrupt the functionality of these applications. Such vulnerability is due to the deployment of VANET in hostile environments, and the cooperative nature of these applications [19]. Context-driven attackers may exploit the aforementioned vulnerabilities and inject false information to influence vehicles' behavior [20]. Consequently, the compromised vehicle can manipulate the data and share false mobility information, which leads to catastrophic accidents that might put many lives in danger. Moreover, the misbehaving vehicle can deceive other benign vehicles to make wrong decisions by responding to illusive events, such as sudden braking and wrong maneuvering, causing serious traffic problems and accidents [11,19]. Moreover, attackers can hijack and control the compromised vehicles and use them to commit many crimes, including terrorist attacks, assassination, and running over crowds of people, to name a few. Several studies have investigated the influences of misbehavior on VANET performance [21–23]. These studies revealed that attackers could significantly degrade the traffic flow, the performance of the routing protocol, and channel utilization. Threats such as artificial collisions [24], illusion attack [25], and Sybil attack [26] are difficult to be detected [27]. The most sophisticated case is when false information resembles real abnormal situation patterns, such as accidents, congestion, and braking patterns [28].

Several solutions have been proposed to defend against insider attacks, namely the context-driven attacks that send false mobility information in VANET. Preventing vehicles from sending false information through cryptographic techniques is computationally expensive and ineffective for many attack scenarios. Meanwhile, many attacks in VANET are inevitable. For example, attackers can manipulate the sensors' data prior to the application of cryptographic techniques. Sensors can also be deceived by simulating the environment around them, making them acquire false data. Hence, vehicles

may spread legitimate but false information [19]. Preventing vehicles from spreading false information in VANET is computationally expensive and could be challenging in many cases [19]. Therefore, misbehavior detection (also called intrusion detection) is crucial for VANET security. Detecting intruders locally and autonomously is the ultimate goal of misbehavior detection in VANET so that an immediate response can be taken. Most of the existing misbehavior detection systems (MDS) employ a set of rules for detecting misbehaving vehicles. These rules assume a stationary environment and identical independence of their input variables. Moreover, several issues have been found in the existing MDS, which adversely affect the performance of the MDS. These issues are related to sensor uncertainty, communication reliability, non-stationary (dynamic) context, and insufficient context representation. However, overlooking sensors' uncertainties, communication reliability, and dynamic vehicular context increase the false alarms rate and decrease the detection rate. The dynamic and heterogeneous noise environment renders data acquisition algorithms unstable, which, in turn, increases the uncertainty of mobility information and adversely affects the information accuracy shared among the vehicles. In addition, vehicles' high mobility and density increase congestion in the communication channels. Such congestion creates unreliable communication and generates incomplete and uncertain mobility data. Nevertheless, extant misbehavior detection solutions have ignored such an issue and used predefined static thresholds in a highly dynamic context instead.

Furthermore, existing solutions are limited to basic and specific attack types. However, many types of attacks can target VANETs, including falsifying vehicle data and hijacking the vehicle's entire system to control a wide range of its functionalities. As such, existing solutions are limited to a particular type of context-driven attacks, namely the basic type; hence, they cannot be generalized. Basic and multifaceted rule sets-based MDS models have been commonly used to detect misbehaving vehicles and intrusions, by building multi-classifiers with independent assumptions. That is, the input variables have been assumed identically and independently distributed (i.i.d.). However, these assumptions are unrealistic for real VANET applications. Such solutions ignore the correlations among the input variables, resulting in a set of non-independent classifiers. Although the output of one classifier might contain information about the output of the other classifiers, which, if utilized, could improve the detection accuracy, such correlation has been ignored by the existing solutions. That is, the correlation among input variables has been overlooked, leading to insufficient context representation, which decreases the detection rate and increases the false alarms rate.

To this end, this paper proposes an ensemble hybrid context-aware misbehavior detection system model (EHCA-MDS), to address the aforementioned two issues, namely, the adaptability of the MDS model with the context, and the insufficient context representation. EHCA-MDS utilizes the concepts of the wisdom of crowds and the power of diversity to improve its effectiveness. Multifaceted adaptive statistical-based classifiers and ensemble learning-based classifiers were integrated to detect the different types of misbehaving vehicles that share false mobility information. More specifically, EHCA-MDS combines several non-parametric, unsupervised-based online statistical classifiers with a supervised-based offline classifier. This combination brings along the robustness and the effectiveness of VANET critical applications. A multifaceted context reference was built and updated online using the Hampel filter, which utilizes the spatial and temporal correlation of the features derived from the mobility information collected from the neighboring vehicles using a Kalman filter-based algorithm. Three types of multifaceted and hybrid features were used to construct the context reference, namely, the consistency-based, plausibility-based, and behavioral-based features. The Kalman filter-based algorithm was used to periodically track and predict the mobility states of the neighboring vehicles due to its efficiency in tracking multiple vehicles' states that are suitable for real-time requirements of VANET applications [29]. During such tracking, the Kalman filter-based algorithm collects recent mobility information of the neighboring vehicles. Using the spatial and temporal properties of the mobility information that were generated from the neighboring vehicles, multidimensional feature sets that represent data plausibility, data consistency, and the behavioral activities were extracted. The Hampel filter was then used to construct a dynamic context-reference using these features. Vehicles

were evaluated from different perspectives using the Hampel filter-based z-score algorithm. The evaluation was based on the vehicle's deviation from the dynamic context reference. As a pre-detection process, multiple adaptive statistical classifiers were proposed using the Hampel filter due to its robustness to outliers and its non-parametric nature that can be built online and fit the dynamic vehicular context. As such, the context reference can be estimated even in the presence of attackers' data. This is because the Hampel filter introduces the breakdown points concepts, in which it considers the smallest percentage of outliers that can render the classifiers biased. The larger the breakdown point an estimator has, the more robust it becomes. This statistical classification model is called the hybrid context-aware misbehavior detection system model (HCA-MDS). HCA-MDS was further improved using the random forest (RF) algorithm to include the correlation among input features and create an ensemble of independent classifiers. The RF was trained using the output coming from the statistical classifiers, along with the context reference model parameters, the consistency features, the plausibility features, and the behavioral features. The random forest algorithm creates a random set of independent classifiers resulting in an unbiased decision. It also selects the most relevant features based on the contribution of each feature that well represents the vehicular mobility data. The final decision was taken based on the output of both statistical and random-forest classifiers. EHCA-MDS is a comprehensive misbehavior detection solution as it combines supervised with unsupervised learning, statistical with machine learning classification, and data-centric with behavioral-centric features in one model, such that it can be generalized for VANET applications.

The rest of this paper is organized as follows. The related works are reviewed in Section 2. The proposed model is elaborated in detail in Section 3. Section 4 presents the performance evaluation and the experimental setup. Section 5 illustrates the results, while Section 6 discusses and analyzes the results. Finally, the paper is concluded in Section 7.

2. Related Work

Due to the tight relation with road safety, business activities, and cybersecurity, VANET is vulnerable to many types of cyberattacks that lead to physical damage, vandalism, terrorism, robbery, kidnapping, and fraud, to name a few. Most of the severe attacks in VANET are context-driven attacks, which manipulate their own context information (mobility data). These attacks might be orchestrated by organizations with a substantial amount of resources, such as government agencies that sponsor cyberwarfare [30]. Such attacks can disturb public safety and/or business activities, leading to substantial loss of lives and properties. Preventing vehicles from sending false information in VANET could be impossible due to the presence of vehicles in hostile environments where the onboard sensors can be manipulated by the owner or technician during maintenance, making vehicles vulnerable to attacks. Although preventing attacks through cryptography-based solutions might be suitable to thwart external attacks, such solutions are not able to prevent the internal attackers that send false information [19]. Because the performance of VANET applications and services depend on exchanging high rate, accurate, and reliable mobility information, attackers that manipulate such information can disrupt the operation of these applications, which put the road safety at risk and adversely affects the traffic efficiency [31]. Likewise, an attacker can employ malware to invade vehicles' operating systems and manipulate the vehicular data, which make the vehicle send false information, such as fake accidents or false congestions messages [19]. Once compromised, the vehicle can be used to carry out many types of crimes, like assassination, hijacking, terrorism, and sabotage. Moreover, malware can force vehicles to respond to fake events, such as hard breaking, which might expose onboard passengers to life threatening accidents. Therefore, misbehavior detection systems (MDSs) are important security measures that work as the second line of defense when the prevention approaches fail. MDSs aim at protecting the integrity of the data and guarantee their correctness.

Misbehavior detection systems (MDSs) have been studied for many years, and several solutions have been suggested [20,32–39]. These solutions can be categorized into three approaches, behavioral [32–38], data-centric [4–44], and hybrid [20,39]. A detailed and recent summary of those

approaches can be found in Reference [30]. The behavioral-based approach has been investigated in intrusion detection and misbehavior detection in VANET. A node would be classified as a misbehaving vehicle if it violated the expected behavior predefined by VANET protocols, applications, and/or services. Moreover, behavioral-based techniques try to encourage cooperation among vehicles. A dedicated watchdog mechanism was adopted by several studies, such as in References [32–35], to monitor the compliance of a vehicle with the predefined expected behavior. The behavioral-based detection starts by monitoring the observable behavior of vehicles, such as forwarding behavior in routing protocols, and broadcasting behavior in broadcasting schemes. For example, monitoring forwarding behavior allows for detecting the wormhole attacks while the broadcasting rate allows for the detection of Denial of Service (DoS) attacks. Although the behavioral approach can effectively detect many types of attacks, it is not suitable for detecting misbehaving vehicles that send false information as it focuses only on monitoring the behavior of the nodes against known protocols and/or services. Vehicles that send false information may not violate those predefined rules or deviate from the expected behavior.

The data-centric approach is feasible to detect misbehaving vehicles that share false information [10,40–44]. Data-centric-based solutions can be further classified into event-based or context-based misbehavior detection. Event-based MDS focuses on detecting false event messages, such as false congestion alerts [44], false crash notifications [38], or false emergency messages [41,45]. However, event-based MDSs are application-specific, as they cover only certain types of events, such as congestion and crash notification. Although a generalization of such an approach has been suggested by many researchers, it is still in the conceptual forms, such as in References [43,46,47], and no implementation has been attempted yet. In addition, the event-based MDS tries to detect if the emerged events were raised due to attack, which is almost difficult to detect locally in ephemeral networks, as it is the case in VANET.

The second type of data-centric-based MDS is context-based MDS. This type of MDS has received much interest from VANET researchers as the performance of most VANET applications and protocols depends on the availability of accurate context information [48]. Therefore, it is commonly believed that context-based data-centric misbehavior detection has many advantages. For example, context-based misbehavior detection can detect most types of attacks that send false information, including false events [29,44,49,50]. Most of the possible attacks in VANET are performed through manipulating context information [51,52]. In addition, attackers in VANET target this type of information as they can inflict major damage easily while they remain undetected [23,25,26,53]. As such, this study aims to build a mobility information-based MDS. Mobility information-based MDSs are more general than the event-based MDSs for locally detecting the misbehaving vehicles that send false information messages. In contrast to event-based MDSs, mobility information-based MDSs try to detect the attacks in the early stages, i.e., before it develops to advanced stages. In other words, the attacks can be detected before the attacker succeeds in triggering the vehicles to send false event messages.

Plausibility and consistency checks are common detection techniques used by existing data-centric solutions. Plausibility-based detection uses a particular known data model of the real-world for detecting implausible information [20,54]. For example, two vehicles cannot occupy the same area at the same time. Similarly, it is implausible for a vehicle to exist in different locations at the same time. Meanwhile, data consistency is used to compare messages received from many independent sources to detect any inconsistencies [10,54]. Unlike message plausibility, data consistency correlates messages originated from different sources. For example, the speed reported by the positioning sensors should be consistent with the vehicle velocity obtained from the speedometer. Plausibility- and consistency-based solutions [10,29,50,51,55–66] can effectively detect many types of attacks under controlled conditions, in which accurate and reliable information can be collected. However, the vehicular environment is highly dynamic, and the collection of accurate and reliable information cannot be guaranteed. Unfortunately, the existing MDS solutions use pre-defined static thresholds to detect the misbehaving vehicles, such that a vehicle whose data consistency or data plausibility exceeds certain thresholds is deemed as misbehaving. However, in the vehicular network where

vehicles move in a harsh environment, communication is unreliable due to the heterogeneous and dynamic noises. Consequently, the data collected are inaccurate and incomplete as a result of message loss that increases due to vehicle mobility and density. As such, static thresholding-based solutions produce high false alarms and low detection rate. On the other hand, a data-centric context-aware MDS (DCA-MDS) proposed by the authors of this paper significantly reduced the false alarms rate and improved the detection accuracy. The solution was successfully able to classify the mobility messages. However, it was not able to identify misbehaving vehicles. In addition, DCA-MDS is vulnerable to context-aware attackers that incrementally manipulate the mobility data and can bypass the dynamic thresholds. In the current study, the DCA-MDS has been modified to identify misbehaving vehicles instead of false messages.

One possible solution to address this issue is by combining the data-centric-based technique with behavioral techniques [67]. Such a combination is referred to as a hybrid-based approach in this paper. The hypothesis is that a context-aware attacker will be keen to convey the false pattern to victims, e.g., by increasing the broadcasting rate, resulting in a different behavior compared to the neighboring vehicles falling in the same context. Bissmeyer and Michael [20] proposed a model that includes behavioral activities in the evaluation of the vehicle. However, predefined static thresholds were used in the detection. In addition, no deep analysis or discussion has been performed to validate the proposed model. In addition, the correlation among the independent classifiers was not considered, which makes the scheme vulnerable to sophisticated attacks, such as advanced illusion attacks where the attackers create traffic illusions by sending consistent and plausible information. Grover and Laxmi [39] have trained MDS models based on a set of behavioral and data-centric features using different machine learning techniques. Random forest generated the highest accuracy. However, no details have been provided about the types of the implemented attacks and the scenario used to generate the training dataset. In addition, the context-features have not been considered in this model, which renders such a solution specific to a particular scenario. Accordingly, Grover and Laxmi's [39] model cannot be generalized. Moreover, the attackers can leverage the model to search for a successful attack. A more general model was proposed in Reference [68], where the datasets were generated under different communication status and environmental noises scenarios. The context was represented by behavioral and data-centric features that were extracted offline, assuming the availability of sufficient information. The MDS model was trained using the neural network algorithm and tested against basic attacks. The derived features were a combination of the consistency features generated using the Kalman filter algorithm and plausibility features generated using the overlapping test algorithm. The results showed the advantage of the proposed model compared to state-of-the-art. However, the model was not evaluated against the context-aware and sophisticated attacks. Ghaleb and Maarof [67] proposed hybrid misbehavior detection that combines both behavioral-based classifiers with data-centric features. However, such an approach has overlooked the correlation between those multifaceted features causing a high false alarms rate.

To protect the VANET from a wide range of attacks, multifaceted, data-centric-based MDSs that utilize data-consistency and plausibility-based rules were commonly suggested by many studies in the literature. However, most such studies rely on predefined static thresholds that do not fit the dynamic nature of VANETs. To improve detection accuracy, hybrid MDSs that combine behavioral-based features with the data-centric-based features were suggested by several other works. However, such an approach has not been deeply studied yet. Most of the solutions in this regard use independent multifaceted rules that ignore the correlation between these rules causing low detection accuracy and high false alarms. The machine learning-based solutions can perceive such relationships. However, existing machine learning-based solutions overlook the context dynamicity and the sophisticated attack types, which render using the supervised machine learning-based solution alone ineffective for VANET dynamic context.

In this paper, a more comprehensive approach called an ensemble-based hybrid context-aware misbehavior detection model for vehicular ad hoc network (EHCA-MDS) was proposed. The wisdom

of crowds and power of diversity concepts have motivated the authors of this paper to construct an ensemble of multifaceted hybrid- and context-aware-based classifiers. Both statistical and machine learning-based approaches were used to build those classifiers. Multidimensional features that reflect the data consistency and plausibility, as well as vehicular behaviors against the communication protocol, were derived to construct the multifaceted statistical-based classifiers. Then, a dynamic context reference model was constructed using Kalman and Hampel filter-based algorithms utilizing the spatial and temporal correlation among mobility information collected from neighboring vehicles. The context reference is dynamically updated according to the newly arrived data from the neighboring vehicles. Vehicles are evaluated using a robust Hampel filter-based outlier z-score algorithm. Vehicles were evaluated based on the deviation from the context reference. The random forest algorithm was used to learn the features of legitimate and misbehaving vehicles, by which a set of diverse classifiers were constructed for the ensemble model. This ensemble was trained based on rich features extracted from the output of the statistical classifiers, the vehicles' data consistency and plausibility, the behavioral scores, and the context model parameters. The outputs of the statistical classifiers and machine learning-based classifiers were aggregated using the weighted voting scheme to deduce the final class.

3. Materials and Methods

The proposed misbehavior detection model is host-based, i.e., it is deployed at each vehicle to detect the local misbehavior in its initial stages before it develops to a sophisticated attack. Due to the absence of labeled attack data, the proposed solution uses an unsupervised statistical method to construct a context reference to detect novel attacks in their early stages. It is a context-aware model, in which a dynamic context reference is constructed and updated online based on the analysis of the spatial and temporal properties of the recent mobility information collected from neighboring vehicles. The proposed model analyzes both the mobility information and the vehicle's cooperative behavior to derive multi-dimensional features that enable detecting a wide range of misbehavior attacks. The mobility information was used because they are the main building block of VANET applications, and they are targeted by many attacks, which have a huge impact on VANET's critical applications and services, as pointed out above. The proposed model is hybrid as it combines a set of multifaceted classifiers to provide diversity that facilitates detecting a vast type of misbehaviors that target VANET's mobility data. Figure 1 shows the architecture of the proposed model.

As shown in Figure 1, the proposed misbehavior detection model consists of six main phases, as follows: firstly, the mobility data collection phase, in which Kalman filter-based algorithms were used to accurately acquire the data from the sensors surrounded by dynamic and heterogeneous noises. The collected data were then efficiently shared among neighboring vehicles. Meanwhile, the data sent by neighboring vehicles were effectively collected. A more detailed description of this phase is presented in Section 3.1. The second phase is the features derivation phase, in which data-centric (namely data-consistency, and data-plausibility) and behavioral features were derived. The detailed description of the derived features is presented in Section 3.2. The third phase is the context representation phase. In this phase, a dynamic context reference is built using an adaptive Hampel filter-based method, utilizing the spatial and temporal properties of the derived features. The output of this phase is the context reference model parameters. Section 3.3. presents a detailed description of this phase. The fourth phase is the multifaceted vehicles evaluation, in which vehicles were evaluated with respect to the derived multidimensional features based on their deviation from the dynamic context reference using the Hampel filter z-score-based algorithm. A set of 13 different classifiers were built (one classifier for each feature) using the Hampel filter-based method, in which vehicles were classified into either misbehaving or benign vehicles. The outputs of this phase are the set of multifaceted, context-aware, statistical-based classifiers, which are used for the pre-detection and the vehicles' scoring. More details about this phase are presented in Section 3.4. In the fifth phase, the outputs of the second phase (the derived consistency, plausibility, and behavioral features), the third phase (context reference model

parameters), and the fourth phase (pre-classification and vehicle scores) were used as input features to train the ensemble learning-based classifiers. The random forest algorithm was used to create a set of independent classifiers. More details about this phase are presented in Section 3.5. In the sixth phase, a weighted average-based method was used to aggregate the output of the classifiers in phase 5 and produce the final decision. More details about this phase are presented in Section 3.6. The detailed description of the used methods is presented in the following subsections.

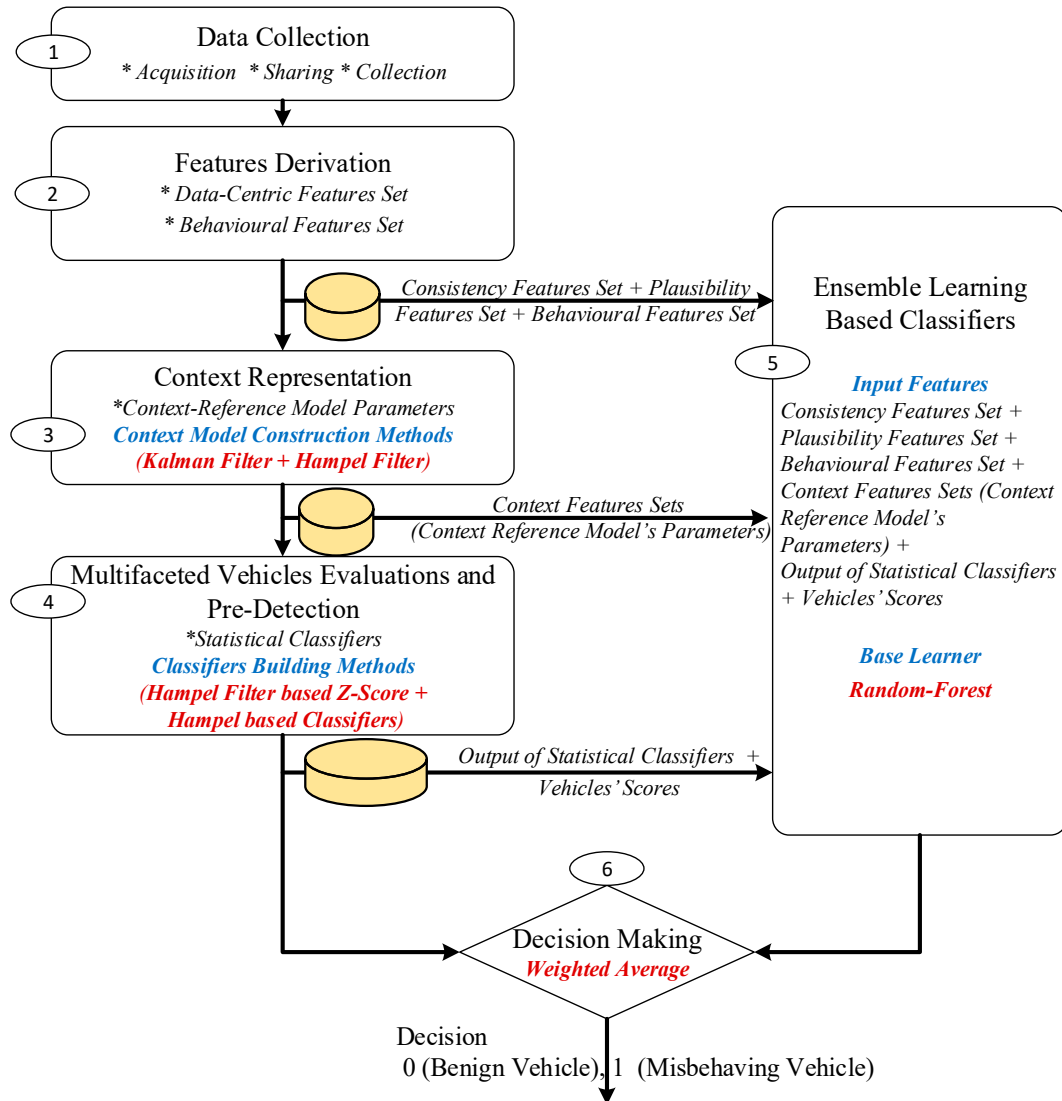


Figure 1. The proposed misbehavior detection system model (ensemble-based hybrid context-aware misbehavior detection system (EHCA-MDS)).

3.1. Data Collection Phase

In this phase, each vehicle collects the mobility information shared by its neighboring vehicles. Data collection in VANET is not a trivial task due to the surrounding noises and harsh communication environment. In particular, vehicles' sensors are surrounded by different types of noises, which render obtaining accurate data challenging. Furthermore, the communication channels between the vehicles are unreliable due to the high mobility, traffic density, and drivers' behavior. Figure 2 shows the data collection scheme used by each node (vehicle or RSU) that hosts the MDS to collect the mobility information. A similar data collection scheme was used recently in our previous publications [67,69]. However, this scheme will be elaborated in more detail.

The scheme in Figure 2 consists of three main components: data acquisition, data-sharing, and data collection. During data acquisition, each vehicle uses the acquisition algorithm to acquire the mobility data recorded by its onboard sensors, such as Global Positioning System (GPS) receiver, speedometer, accelerometer, and gyroscope. However, the harsh environment that these vehicles operate within, as well as the mobility nature of those vehicles, contained heterogeneous and dynamic noises within the data, which render the acquired data uncertain and the environment-dependent, i.e., it depends on the environment surrounding the vehicles' sensors. A robust acquisition algorithm that is aware of such an environment should be used to improve the accuracy and estimate the uncertainty of the acquired data. Although there are many acquisition algorithms in the literature, the improved innovation-based adaptive estimation Kalman filter algorithm proposed in Reference [70] has been found robust for heterogeneous and dynamic noises in vehicular environments. The algorithm in Reference [70], which is called Enhanced Innovation-based Adaptive Estimation Kalman Filter algorithm (EIAE-KF), estimates the measurement noise covariance according to the discrepancy between the prediction and measurement phases of the Kalman filter algorithm in a timely manner. The estimated noise covariance was then used to adaptively adjust the Kalman gain to improve the estimation accuracy along with estimating the uncertainty of the data. For more details about the EIAE-KF algorithm, the readers are referred to the study in Reference [70].

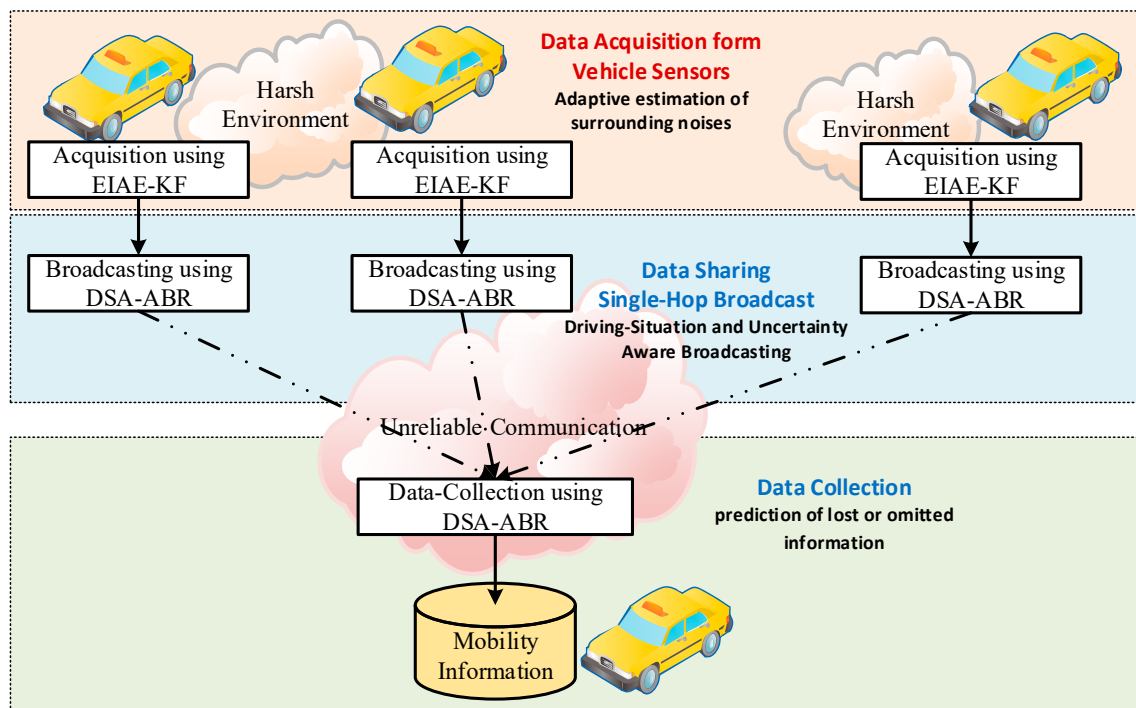


Figure 2. Data collection scheme.

During the data sharing step, each vehicle shares its mobility information messages with its neighbors. Due to the high mobility of vehicles, mobility data become outdated quickly. Therefore, vehicles should broadcast their mobility information at high rates (10 messages per second according to the VANET standards) [71]. Unfortunately, the high broadcasting rate in shared communication channels increases the congestion rate within these channels, which in turn increases the rate of lost messages. Moreover, intermittent communication increases the uncertainty of the information, leading to degrading the effectiveness of the applications and increases the false alarm rate in the existing Intrusion Detection Systems (IDSs) solutions. Many broadcasting schemes have been proposed in the literature to adapt the broadcasting rate according to channel characteristics such as busy time status, vehicle density, traffic flow, or a combination of these parameters. In this study, the

Driving-Situation-Aware Adaptive Broadcasting Rate Scheme (DSA-ABR) [31], has been adopted to broadcast the mobility information. DSA-ABR adapts the broadcasting rate of each individual vehicle according to the driving situation to decrease the number of broadcasted messages, and thus decreases the number of competitors on the channel while preserving the accuracy of the information. Furthermore, DSA-ABR increases broadcasting according to the changes in driving behavior, such as during braking, acceleration, and maneuvering, and it decreases the broadcasting rate when the driving regime is stable and/or mobility information has high uncertainty. During the stable driving situation, neighboring vehicles can accurately predict the omitted mobility messages. On the other hand, if the uncertainty of the information is high, there is no need to increase the broadcasting rate.

For a reliable collection of the mobility information broadcasted by neighboring vehicles, the DSA-ABR uses a Kalman filter prediction algorithm to track the mobility information broadcasted by neighboring vehicles and predict the lost or omitted information. It is worth noting that some mobility information messages are lost due to unreliable communication in the vehicles' environment. In the DSA-ABR scheme, some mobility information messages that can be easily predicted by receivers are deliberately omitted by senders to decrease the broadcasting rate, so as to improve communication reliability. The output of this phase is a dataset for each neighboring vehicle. Each dataset contains a history of vehicle mobility information sampled at every 100 milliseconds.

3.2. Features Derivation Phase

After collecting the mobility information of the neighboring vehicles, two main types of features were derived: a data-centric-based features set and a behavioral-based features set.

3.2.1. Data-Centric-Based Features

Two data-centric types of features were extracted from each dataset collected in the previous phase: a consistency-based features set and a plausibility-based features set. More specifically, the vector of the innovation errors of the Kalman filter algorithm is used as consistency features. This vector contains the divergence between the predicted state and the received mobility messages. Thus, four consistency features related to the positioning error, and speed prediction errors in both latitude and longitude dimensions, were extracted, assuming two-dimensional space and a fixed acceleration mobility model. The plausibility-based features set consists of three main features, as follows. The communication range-based feature, which is the distance between the sender (all neighboring vehicles) and the receiver vehicle (the subject vehicle that host the misbehavior detection model). The distance of the vehicles' appearance feature, which is the distance between the sender and the receiver vehicle when the sender vehicle enters the communication range of the receiver vehicle, and the overlapping frequency-based feature, which is the number of times in which the vehicle occupation area of the neighboring vehicles is overlapped with each other. Thus, seven data-centric-based features were derived to represent the data consistency (DC1, DC2, DC3, and DC4) and the plausibility (DC5, DC6, and DC7). Table 1 presents a list of those features with their names and description. In the table, DSF stands for Data-centric Features Set while DC stands for Data-Centric.

The following paragraphs explain more about how the consistency features (DC1, DC2, DC3, and DC4) were derived. As the innovation sequence of the Kalman filter can describe the discrepancy between the expected and reported mobility information, it has been used to represent the consistency of the received information from neighboring vehicles with the information predicted using the mobility model. Accordingly, there are two main steps to derive the consistency score (CS) feature. Where $e_k^{v_i}$ is a vector that contains the innovation errors of vehicle v_i that was recorded at the time epoch (k), and E is a matrix that contains a list of the innovation errors for all neighboring vehicles. The consistency feature $f_{e(k)}^{v_i}$ of a neighboring vehicle v_i at a specific time epoch, k is then calculated according to the following equation:

$$\forall e_k^{v_i} \in E : f_{e(k)}^{v_i} = \begin{cases} y_k^{v_i} - \tilde{y}_{k|k-\tau(i)}^{v_i} & \text{if a message is received} \\ f_{k-\tau(i)}^{v_i} * (k - \tau(i)) & \text{Otherwise} \end{cases} \quad (1)$$

where $y_k^{v_i}$ is the mobility information vector as received from the neighboring vehicle v_i at the time epoch k , $\tilde{y}_{k|k-\tau(i)}^{v_i}$ is the predicted mobility information vector using the last received messages (at the time epoch $\tau(i)$), and $\tau(i)$ is the time epoch when the last message is received.

Table 1. The derived data-centric-based features (consistency and plausibility).

DFS#	Type	Name	Description
DC1	f_1	Consistency	Latitude prediction error
DC2	f_2		Longitude prediction error
DC3	f_3		Latitude speed Prediction error
DC4	f_4		Longitude speed prediction error
DC5	f_5	Plausibility	Communication Range
DC6	f_6		Vehicle appearance distance
DC7	f_7		Overlapping Frequency

The plausibility-based features were driven as follows. For the communication range-based feature (denoted by DC5 and f_5 in Table 1), v contains the list of all neighboring vehicles, $p(x, y)$ is the position of the vehicle that will drive this feature (the receiver), and $p_i(x_i, y_i)$ is the position of the vehicle v_i that sent the mobility messages. Then, the communication range-based feature $f_{cr(k)}^{v_i}$ of vehicle v_i at time epoch k can be calculated as Euclidean distance between the sender and receiver, as follows:

$$\forall v_i \in v : f_{cr(k)}^{v_i} = \forall v_i \in v : f_{cr(k)}^{v_i} = \|p(x, y) - p_i(x_i, y_i)\| \quad (2)$$

The vehicle appearance distance (DC6 or f_6) is calculated for newly neighboring vehicles when they enter the communication range of the subject vehicle (the receiver). Thus, the $f_{va(k)}^{v_i}$ can be set equal to $f_{cr(k)}^{v_i}$, assuming k is the time epoch when v_i enter the communication range of the subject vehicle. The overlapping frequency feature (denoted by DC7 or f_7 in Table 1) was derived based on the work of Bissmeyer et al. in Reference [20]. In their work, the vehicles were modeled as rectangles so that vehicles are misbehaving if their rectangles intersect (overlap) with each other. This study utilized the same idea to derive overlapping-based features that work as an indication of possible misbehavior. The algorithm in Figure 3 presents a pseudocode for the data-centric-based features derivation.

Algorithm 1: Ensemble of Multifaceted Classifiers**Input:** $M_{(k)}$ mobility messages with vehicles' dimensions width (w) and length (l)**Output:** $O_{f(k)}^{v_i}$ Classification Output, $S_{(k)}$ Scores, $CRM_{(k)}$ Context Reference Model**1: FOR** each time epoch k **DO** Use Kalman Filter to **Derive** the consistency features from $f_{e(k)}^{v_i}$ as follow
$$2: \quad \forall e_k^{v_i} \in E: f_{e(k)}^{v_i} = \begin{cases} y_k^{v_i} - \check{y}_{k|k-\tau(i)}^{v_i} & \text{if a message is recieved} \\ f_{k-\tau(i)}^{v_i} * (k - \tau(i)) & \text{Otherwise} \end{cases}$$
 Derive the communication range based plusablity feature (d)3: $\forall v_i \in v: f_{cr(k)}^{v_i} = \|p(x, y) - p_i(x_i, y_i)\|$ Derive the Vehicle appearance distance based plausibility feature (d) for each vehicle i in its vicinity4: $\forall v_i \notin v: f_{va(k)}^{v_i} = f_{cr(k)}^{v_i}$ for newly appeared vehicles5: **Derive** the overlap based plausibility features (d) for each vehicle i in its vicinity6: $\forall v_i \text{ and } v_j \in v \text{ find } d_{(i,j)}: d_{(i,j)} = \|p_i(x_i, y_i) - p_j(x_j, y_j)\|$ 7: Calculate $d_{min(i,j)} = \min\left(\left(\frac{w_i + w_j}{2}\right), \left(\frac{l_i + l_j}{2}\right)\right)$ 8: $\forall v_i \text{ and } v_j \in v \text{ if } |d_{(i,j)}| < d_{min(i,j)} \text{ then}$ **Set** $f_{ol(k)}^{v_i} = f_{ol(k)}^{v_i} + 1$ and **SET** $f_{ol(k)}^{v_j} = f_{ol(k)}^{v_j} + 1$ **8: END FOR LOOP****Figure 3.** Pseudocode for features derivation.**3.2.2. Behavioral-Based Features Set**

As shown in Table 2, vehicles' behaviors are represented by the following six features. Four features (BF1, BF2, BF3, and BF4) were attributed to the vehicle's behavior against the communication protocol, while two features (BF5 and BF6) are for driving behavior features.

Table 2. The derived behavioral-based features.

BF#	Feature Name		Description
BF1	f_8	Connection Length	The time difference between the first received mobility message from each neighboring vehicle and the current time epoch.
BF2	f_9	Received Messages	The total number of messages that are received from each neighbor.
BF3	f_{10}	Broadcasting Rate	The moving average of the number of received messages divided by the connection length.
BF4	f_{11}	Broadcasting delay	The moving average of the differences between time of receiving the messages by the neighboring vehicles and their creation time in the subject vehicle.
BF5	f_{12}	Jerk acceleration	The rate of change of acceleration
BF6	f_{13}	Speed deviation	The divergence between the sender's average speed and the median speed of all neighboring vehicles

3.3. Context Representation Phase

In this phase, the context reference model is constructed and updated in a timely manner. Because neighboring vehicles are exposed to the same context in terms of traffic conditions, environmental noises, communication status, and road status, thus, the spatial correlation among the features derived in the previous phase can be utilized to construct a dynamic context reference. The spatial correlation among neighboring vehicles has been utilized to construct the context reference. The Hampel filter has been used to construct the reference model as follows: let $F = \{f_1, f_2, f_3, \dots, f_m\}$ is the set of features, $V = \{v_1, v_2, v_3, \dots, v_n\}$ is the set of neighboring vehicles at time epoch k . Also, X_k is a matrix that

contains the value of the features of each neighboring vehicle at time epoch k , $x_k^{NV(i)}$ is a vector that contains the values of feature $j \in m$ derived from each neighboring vehicle $v_i \in V$ at time epoch k .

$$X_{(k)} = \begin{bmatrix} x_{(k)}^{v_1} \\ x_{(k)}^{v_2} \\ \vdots \\ x_{(k)}^{v_n} \end{bmatrix} = \begin{bmatrix} f_{1(k)}^{v_1} & f_{2(k)}^{v_1} & \cdots & f_{m(k)}^{v_1} \\ f_{1(k)}^{v_2} & f_{2(k)}^{v_2} & \cdots & f_{m(k)}^{v_2} \\ \vdots & \vdots & \cdots & \vdots \\ f_{1(k)}^{v_n} & f_{2(k)}^{v_n} & \cdots & f_{m(k)}^{v_n} \end{bmatrix}, \forall v_i \in V \text{ and } f_j \in F \quad (3)$$

where $f_{j(k)}^{v_i}$ is the value of the feature j that was derived from the vehicle i at time epoch, k . Thus, the Hampel filter-based dynamic context reference model $CRM_{(k)}$, at time epoch k , can be computed as follows:

$$CRM_{(k)} = \begin{cases} \varnothing_{(k)} = \text{median}(X_{(k)}) \\ \delta_{(k)} = 1.4826 \times \text{median}\{|X_{(k)} - \varnothing_{(k)}|\} \\ CRUB_{(k)} = HUB_k = \varnothing_{(k)} + \beta \times \delta_{(k)} \\ CRLB_{(k)} = HLB_k = \varnothing_{(k)} - \beta \times \delta_{(k)} \end{cases} \quad (4)$$

where $\varnothing_{(k)}$ and $\delta_{(k)}$ are the median and the median absolute deviation (MAD) of each feature in $X_{(k)}$, respectively. $CRUB_{(k)}$ is the context reference's upper bound which is equal to the Hampel filter upper bound (HUB_k). $CRLB_{(k)}$ is the context reference's lower bound which is equal to the Hampel filter lower bound (HLB_k), and β is a threshold that was selected heuristically, such that the best accuracy is obtained. Thus, the model parameters' can be expressed as follows:

$$CRM_{(k)} = \begin{bmatrix} \varnothing_{(k)}^{f_1} & \varnothing_{(k)}^{f_2} & \cdots & \varnothing_{(k)}^{f_m} \\ \delta_{(k)}^{f_1} & \delta_{(k)}^{f_2} & \cdots & \delta_{(k)}^{f_m} \\ CUB_{(k)}^{f_1} & CUB_{(k)}^{f_2} & \cdots & CUB_{(k)}^{f_m} \\ CLB_{(k)}^{f_1} & CLB_{(k)}^{f_2} & \cdots & CLB_{(k)}^{f_m} \end{bmatrix} \quad (5)$$

Because vehicular context is highly dynamic due to high vehicle mobility, critical VANET applications require a high rate of context-awareness messages. Thus, it is worth noting that the context references are built on the fly and updated every 100 ms to capture the highly dynamic temporal change of the context data. However, building the context reference based on information collected within a very short period can make the context reference unstable and inefficient, causing improper detection accuracy and delay. Similarly, using a long period can make the context reference rigid. Accordingly, Welford's approach [73] to calculate the running average is used to update the $CRM_{(k)}$ model parameters in real-time, as follows:

$$CRM_{(k)} = \frac{CRM_{(k-1)} * (w - 1)}{w} + \frac{CRM_{(k)}}{w} = CRM_{(k-1)} + \frac{CRM_{(k)} - CRM_{(k-1)}}{w} \quad (6)$$

where $CRM_{(k)}$ is the context reference model at time epoch k , and w is the sliding window width. Equation (4) is equivalent to the average of the context references using the average batch formula $\sum_{i=k-w}^k \frac{CRM_{(i)}}{w}$, but it is numerically stable and more efficient than the average batch formula.

3.4. Multifaceted Vehicles Evaluation and Pre-Detection Phase

After constructing the context reference, vehicles are evaluated based on their deviation from the context reference. The evaluation has been calculated using a Hampel Filter-based z-score outlier detection ($z = \frac{x_i - \mu}{\sigma}$). Hampel filter replaces the arithmetic mean μ , and standard deviation σ by the

median \varnothing_k and median absolute deviation (MAD) δ_k , respectively. Thus, each vehicle got a score for each feature using Hampel based z-score, which can be rewritten as follows:

$$z_{f_j(k)}^{v_i} = \frac{f_{j(k)}^{v_i} - \varnothing_{(k)}^{f_j}}{\delta_{(k)}^{f_j}}, \forall v_i \in V \text{ and } f_j \in F \quad (7)$$

The set of z-scores of a vehicle, i , at time epoch, k , was represented by a vector, $Z_{(k)}^{v_i}$, as follows:

$$Z_{(k)}^{v_i} = \left[\begin{array}{cccc} \frac{f_{1(k)}^{v_i} - \varnothing_{(k)}^{f_1}}{\delta_{(k)}^{f_1}} & \frac{f_{2(k)}^{v_i} - \varnothing_{(k)}^{f_2}}{\delta_{(k)}^{f_2}} & \cdots & \frac{f_{m(k)}^{v_i} - \varnothing_{(k)}^{f_m}}{\delta_{(k)}^{f_m}} \end{array} \right] \quad (8)$$

Accordingly, the set of all scores of all vehicles can be represented by a matrix, $S_{(k)}$, as follows:

$$S_{(k)} = \left[\begin{array}{c} Z_{(k)}^{v_1} \\ Z_{(k)}^{v_2} \\ \vdots \\ Z_{(k)}^{v_n} \end{array} \right] = \left[\begin{array}{cccc} z_{f_1(k)}^{v_1} & z_{f_2(k)}^{v_1} & \cdots & z_{f_m(k)}^{v_1} \\ z_{f_1(k)}^{v_2} & z_{f_2(k)}^{v_2} & \cdots & z_{f_m(k)}^{v_2} \\ \vdots & \vdots & \cdots & \vdots \\ z_{f_1(k)}^{v_n} & z_{f_2(k)}^{v_n} & \cdots & z_{f_m(k)}^{v_n} \end{array} \right] \quad (9)$$

Different statistical classifiers were developed according to the context reference and vehicle scores. With each vehicle, a set of tests has been conducted for each feature to the prediction results, as follows:

$$o_{f_j(k)}^{v_i} = \begin{cases} 0 & \text{benign vehicle} & \text{if } CLB_{(k)}^{f_j} < z_{f_j(k)}^{v_i} < CUB_{(k)}^{f_j} \\ 1 & \text{misbehaving vehicle} & \text{Otherwise} \end{cases} \quad (10)$$

where $o_{f_j(k)}^{v_i}$ is the output of the classifier that represents the features, f_j , for a vehicle, v_i , at the time epoch, k . $CLB_{(k)}^{f_j}$ and $CUB_{(k)}^{f_j}$ are the context lower bound and the context upper bound respectively, as computed from the feature, f_j , of all vehicles present at time epoch, k , using the Hampel filter. If the $z_{f_j(k)}^{v_i}$ is located outside those bounds, the corresponding vehicle (v_i) is considered as a misbehaving vehicle with respect to the feature, f_j , detected at time epoch, k .

Figure 4 presents the multifaceted MDS model, which contains both the data-centric context-aware MDS (DCA-MDS) and the hybrid context-aware MDS (HCA-MDS). We have distinguished between these two MDSs in order to separately evaluate the performance of each detection concept. The DCA-MDS model is a context-aware MDS that has been built based on the data-centric features. It aims to thwart many types of basic attacks where the attacker is unaware of the context. However, a context-aware or a sophisticated attack type can bypass such a solution rendering DCA-MDS not effective. Thus, the behavioral features-based classifiers were built to assist the DCA-MDS to improve the detection rate. Our hypothesis is that the attackers may show different behavioral activities compared to the benign vehicles because they are keen to convey their false information to the vehicles in their vicinity. Thus, the hybrid model, HCA-MDS, is constructed. The model considers both data-centric- and behavioral-based features. Figure 5 shows the pseudocode for features derivation, context reference model construction, and the set of classification rules that were used by HCA-MDS and DCA-MDS. Table 3 lists the description of the symbols used in Figure 5.

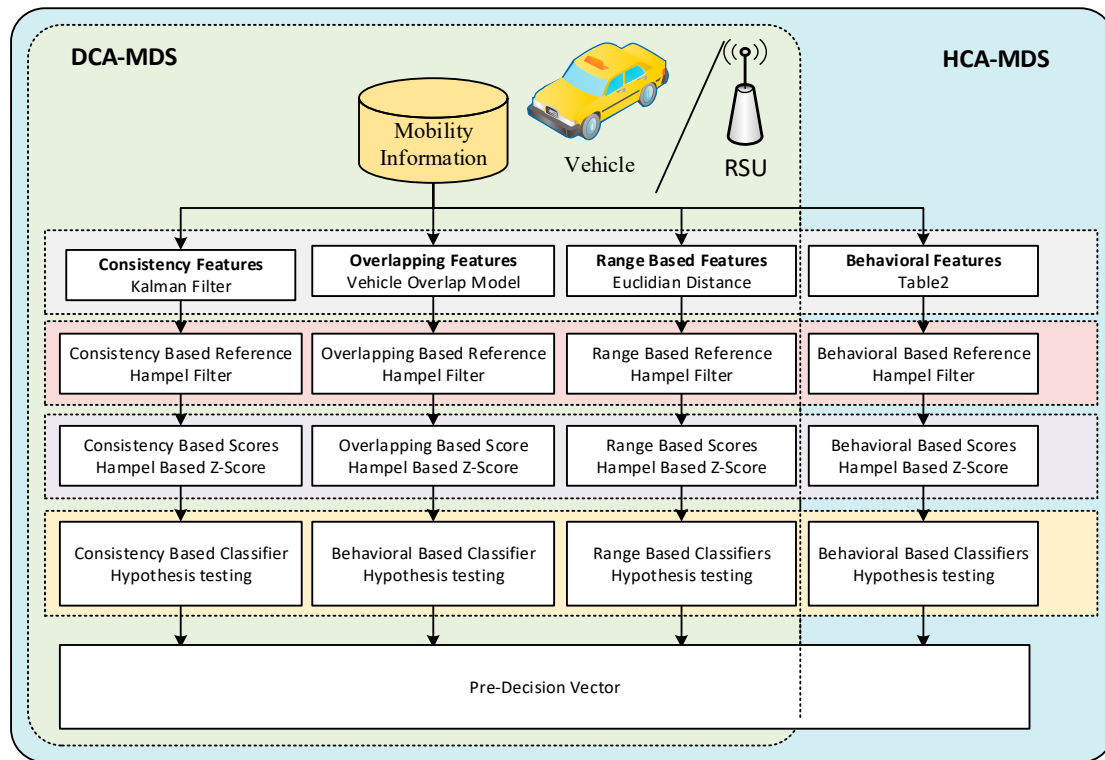


Figure 4. Hybrid multifaceted MDS model (HCA-MDS and DCA-MDS).

Table 3. Symbols Description.

Symbol	Description
$f_{e(k)}^{v_i}$	Consistency features: innovation errors of vehicle (v_i) at time epoch (k)
$y_k^{v_i}$	Received mobility data
$\hat{y}_{k k-\tau(i)}^{v_i}$	Predicted mobility data by Kalman filter
$d_{(i,j)}$	The distance between the vehicle v_i at position $p_i(x_i, y_i)$ and vehicle v_j at position $p_j(x_j, y_j)$
$\tau(i)$	The time epoch of last received mobility data
$f_{cr(k)}^{v_i}$	Range-based features: the distance between the current vehicle and the neighboring vehicle
$os_k^{v_i}$	Overlapping-based feature
$d_{min(i,j)}$	The minimum accepted distance between vehicles v_i and v_j
$f_{j(k)}^{v_i}$	Feature number j for vehicle v_i at time epoch k
$f_{b(k)}^{v_i}$	Behavioral features of the vehicle v_i at time epoch k
$CRM_{(k)}$	Context-reference model parameters
\varnothing_k	Median of the vehicles' temporal summaries
δ_k	The median absolute deviation of vehicles' temporal summaries
HUB_k	Hampel upper bound, also called $CRUB_{(k)}$ context reference upper bound
HLB_k	Hampel lower bound, also called $CRLB_{(k)}$ context-reference lower bound
β	Tuning parameters
$z_{f_j(k)}^{v_i}$	Vehicle v_i Hampel-based Z-score with respect to feature $f_{j(k)}^{v_i}$ at time epoch k
$\sigma_{f_j(k)}^{v_i}$	Classification rule

Algorithm 2: Ensemble of Multifaceted Classifiers**Input:** $M_{(k)}$ mobility messages with vehicles' dimensions width (w) and length (l)**Output:** $O_{f(k)}^{v_i}$ Classification Output, $S_{(k)}$ Scores, $CRM_{(k)}$ Context Reference Model**1: FOR** each time epoch k **DO** Use Kalman Filter to **Derive** the consistency features from $f_{e(k)}^{v_i}$ as follow

$$2: \quad \forall e_k^{v_i} \in E: f_{e(k)}^{v_i} = \begin{cases} y_k^{v_i} - \check{y}_{k|k-\tau(i)}^{v_i} & \text{if a message is recieved} \\ f_{k-\tau(i)}^{v_i} * (k - \tau(i)) & \text{Otherwise} \end{cases}$$

Derive the communication range based plusability feature (d)

$$3: \quad \forall v_i \in v: f_{cr(k)}^{v_i} = \|p(x, y) - p_i(x_i, y_i)\|$$

 Derive the Vehicle appearance distance based plausibility feature (d) for each vehicle i in its vicinity

$$4: \quad \forall v_i \notin v: f_{va(k)}^{v_i} = f_{cr(k)}^{v_i} \quad \text{for newly appeared vehicles}$$

Derive the overlap based plausibility features (d) for each vehicle i in its vicinity

$$6: \quad \forall v_i \text{ and } v_j \in v \text{ find } d_{(i,j)}: d_{(i,j)} = \|p_i(x_i, y_i) - p_j(x_j, y_j)\|$$

$$7: \quad \text{Calculate } d_{min(i,j)} = \min\left(\left(\frac{w_i + w_j}{2}\right), \left(\frac{l_i + l_j}{2}\right)\right)$$

$$8: \quad \forall v_i \text{ and } v_j \in v \text{ if } |d_{(i,j)}| < d_{min(i,j)} \text{ then}$$

$$\quad \text{Set } f_{ol(k)}^{v_i} = f_{ol(k)}^{v_i} + 1 \text{ and SET } f_{ol(k)}^{v_j} = f_{ol(k)}^{v_j} + 1$$

Drive the behavioral features according to the Table 2 Concatenate the feature matrices of all vehicles $f_{b(k)}^{v_i}$ **Apply- Hampel-Filter** to generate the context reference model CRM_k .

$$10: \quad \text{Compute } CRM_{(k)} = \begin{cases} \emptyset_{(k)} = \text{median}(X_{(k)}) \\ \delta_{(k)} = 1.4826 \times \text{median}\{|X_{(k)} - \emptyset_{(k)}|\} \\ CRUB_{(k)} = HUB_k = \emptyset_{(k)} + \beta \times \delta_{(k)} \\ CRLB_{(k)} = HLB_k = \emptyset_{(k)} - \beta \times \delta_{(k)} \end{cases}$$

 $CRM_{(k)}$ is the consistency based context reference model at the time epoch k . **Update the context reference**

$$11: \quad CRM_{(k)} = CRM_{(k-1)} + \frac{CRM_{(k)} - CRM_{(k-1)}}{\text{time_windwo}}$$

$$12: \quad \text{Compute vehicle score } z_{f_j(k)}^{v_i} = \frac{f_{j(k)}^{v_i} - \emptyset_{(k)}^{f_j}}{\delta_{(k)}^{f_j}}, \forall v_i \in V \text{ and } f_j \in F$$

$$13: \quad \text{Detect Misbehaving Nodes } o_{f_j(k)}^{v_i} = \begin{cases} 0 & \text{benign vehicle if } CRLB_{(k)}^{f_j} < z_{f_j(k)}^{v_i} < CRUB_{(k)}^{f_j} \\ 1 & \text{misbehaving vehicle Otherwise} \end{cases}$$

8: END FOR LOOP**Figure 5.** Pseudocode for features derivation, context reference model construction, and the set of classification rules (HCA-MDS and DCA-MDS).**3.5. Ensemble Learning Phase**

In the previous section, the vehicles were evaluated, and pre-detection was obtained. Multi-faceted statistical adaptive classifiers were built, as shown in Figure 3. This design relies on the univariate statistical analysis, in which the input variables (the derived features) were assumed to be independent. That is, a classification rule was created for each derived feature. The relationship among the variables has been ignored, rendering such design vulnerable to sophisticated and context-aware attackers. An attacker who incrementally manipulates the mobility information can end up with a successful attack because the independent classifiers can only distinguish the outliers that deviate much from the statistical context reference represented by that specific feature. Therefore, considering the relationship between the input features is important to detect such advanced attacks. For example, the broadcasting rate feature, which is a behavioral-based feature, is correlated with the consistency features (Kalman innovation errors). If the broadcasting rate increased, the innovation errors of the Kalman filter would decrease. An attacker who incrementally manipulates the vehicle's own mobility information will be

keen to convey the false information to the neighboring vehicle by increasing the broadcasting rate. The attacker will also be aware of the context, so that the increase of the broadcasting rate will be within the context reference boundaries. Thus, independent classifiers that ignore the correlation between the features may not be able to detect such an attack. Furthermore, collecting sufficient information about the vehicles can be challenging in the VANET environment. Therefore, mining previously known data can be useful in detecting previously seen patterns. Because correlation among the features is stationary with respect to the particular vehicular context, a supervised approach can be acceptable to predict the sophisticated attack context-aware attackers. Accordingly, an ensemble-learning-based predictive model has been constructed using the random forest algorithm. A random forests algorithm creates a forest of random decision trees-based classifiers. Each of those classifiers was trained based on a random sample that was randomly selected with replacement from the original dataset. The outputs of the classifiers are aggregated using the majority voting scheme. Inspired by the concepts “wisdom of crowds and the power of diversity,” the random forest algorithm was selected to improve the detection effectiveness and provide comprehensive context representation. Figure 6 shows the procedure used to train and evaluate the proposed predictive model.

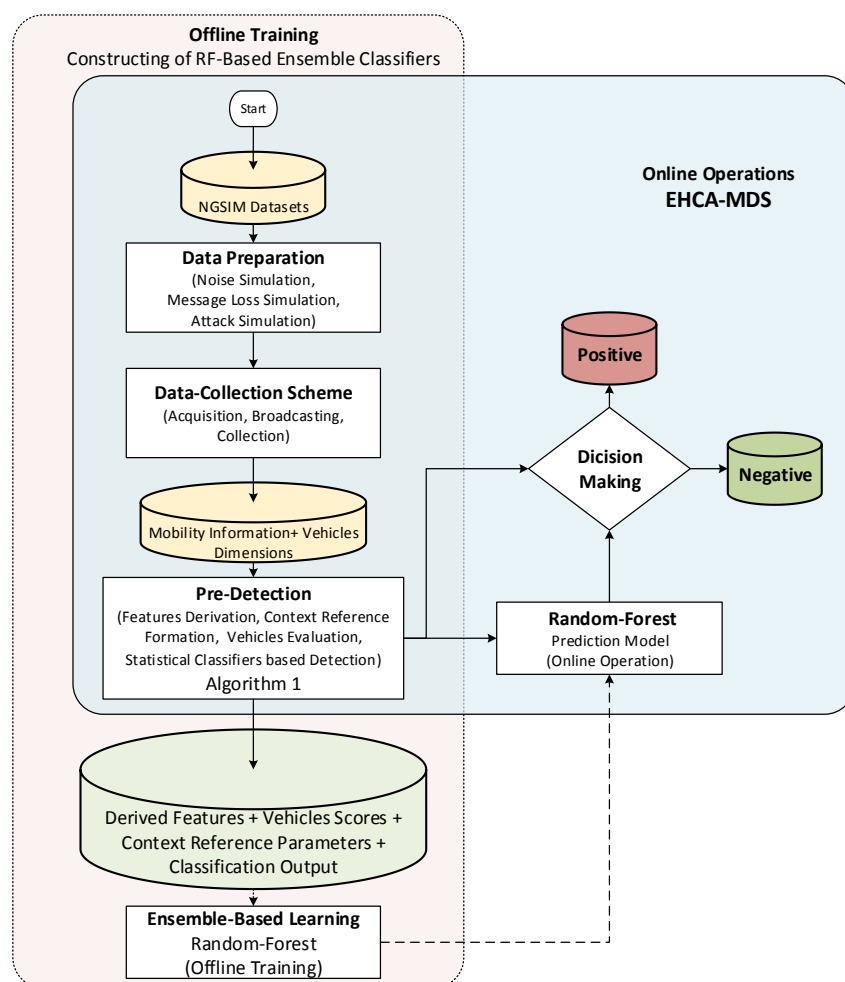


Figure 6. Flowchart of the ensemble model construction.

As shown in Figure 6, there are two main tasks conducted to construct the proposed EHCA-MDS model. The offline training of the random forest-based classifiers and the integration of the trained ensemble classifiers with the HCA-MDS model presented in Figure 4 for the online operation. For offline training, the first procedure is the preparation of the datasets. The preparation includes noise injection to simulate the vehicular noise environment. The datasets have been replayed under different

communication reliability scenarios, where different message loss ratios were simulated. Then, many basic and sophisticated attack types related to false mobility information attacks have been simulated and injected into vehicle trajectories. A detailed description of the procedure of the data preparation is presented in the next section (Section 4). Then, the HCA-MDS model is implemented and used by each vehicle in the datasets to detect the simulated misbehaving vehicles. Each vehicle uses the proposed algorithm shown in Figure 5 to derive the context features, construct the context reference model, evaluate the vehicles, and provide pre-detection outputs. Then, the derived features, the parameters of the context reference model, the vehicle's scores, and the classification outputs have been used as input features to the random forest algorithm to construct the ensemble-based classifiers. The integration of the trained ensemble classifiers with the HCA-MDS model was then carried out for the online operations. The classification outputs of all classifiers (the statistical and the ensemble-based classifiers) were aggregated using a weighted average for the final decision.

3.6. Decision Phase

After obtaining the outputs of the statistical and ensemble learning-based classifiers, the final decision is taken using the following formula:

$$D_{(k)}^{v_i} = \begin{cases} 0 & \text{benign vehicle if } \frac{\sum w_j o_j}{\sum w_j} > \min(W) : w_j \in W \\ 1 & \text{misbehaving vehicle Otherwise} \end{cases} \quad (11)$$

where $D_{(k)}^{v_i}$ is the final decision, and w_i and o_i are the weight and the output of the classifier j , respectively. The classifiers' weight vector, W , is obtained from the performance of the classifiers according to the performance of the individual classifiers. For example, if the performance of the classifier j in terms of F-measure is 80%, the classifier is then weighted as 0.8. Consequently, if the condition in Equation (9) was true, the misbehaving vehicle would be detected.

4. Performance Evaluation

The experimental setup and performance metrics are discussed in the following subsections.

4.1. Experimental Setup

In this study, the next generation simulation (NGSIM) dataset, a real-world traffic dataset that contains vehicle trajectories recorded each 100 ms, was used to conduct the experiments. It has been generated by the Federal Highway Administration (FHWA) of the United States Department of Transportation (DOT) [74]. It represents the ground truth information of neighboring vehicles' trajectories [74,75]. Extensive simulations have been conducted to evaluate the proposed model. In particular, many activities have been done, such as dataset preparation, environmental noise injection, communication simulation, and misbehavior simulation. Similar to related studies [2,20,29,35,76,77], the Matlab tool [78] has been used for simulating the environmental noises, vehicle communication, and misbehaving vehicles.

4.1.1. Sample Selection

The dataset is categorized into four different clusters, based on driver behavior to ensure that the evaluation covers all types of driver behaviors. For each vehicle, three features were selected to represent driver behavior, namely: time headway, space headway, and lane changing ratio. The selected features are aggregated by finding their mean and variance, which were then used as input for the K-means clustering algorithm [79]. These clusters describe four types of driving regime: free-flowing, random flowing, car flowing, and lane changing behavior. The purpose of this categorization is to ensure that the vehicle behaviors have no influence on the performance of the proposed scheme. Table 4 shows the dataset samples used to evaluate and validate the proposed EHCA-MDS scheme.

As shown in the table, 16 datasets have been used for the evaluation. Those datasets are composed of Cooperative Awareness Messages (CAMs) collected by 15 vehicles, as well as one simulated RSU. Those vehicles were selected randomly from different driving regimes and used to host the evaluated misbehavior detection models.

Table 4. Next generation simulation (NGSIM) dataset selected samples (host vehicles).

Dataset	Host Vehicle Id	Vehicle Regime	Average Speed (m/s)	Duration (s)	Dataset Size CAMs	Total Neighbors
DS1	13	Free-Flow	16.8	94.8	113,258	177
DS2	252	Free-Flow	23.3	55.8	155,908	255
DS3	455	Free-Flow	21.6	60.3	145,904	260
DS4	2280	Free-Flow	24.0	54.1	197,511	270
DS5	5	Lane-Change	22.4	70.2	80,568	119
DS6	1133	Lane-Change	31.7	39.3	107,565	214
DS7	1687	Lane-Change	22.4	76.5	110,051	314
DS8	1	Lane-Change	18.0	88.4	88,971	134
DS9	268	Flowing-Mode	26.2	58.5	156,941	255
DS10	1066	Flowing-Mode	33.0	47.5	111,305	225
DS11	1964	Flowing-Mode	21.5	72.9	223,211	317
DS12	7	Flowing-Mode	22.5	71.1	85,244	127
DS13	1593	Flowing-Mode	21.1	74.2	186,260	294
DS14	2885	Random-Flow	16.5	94.5	150,127	200
DS15	1899	Random-Flow	19.9	78.8	231,867	331
DS16	RSU	Mixed	28.0	57.3	479,823	284

4.1.2. Simulation of Environmental Noises

Various types of environmental noises have been injected into the vehicle trajectories in the NGSIM data to represent VANETs' harsh environment. A combination of stationary white noises with zero mean, non-stationary white noise with time-varying variance, and correlated noises have been reported by many studies in VANET context acquisition and used to simulate the dynamic and heterogeneous environmental noise in VANET [18,70,80,81]. Noise injection is a common procedure to simulate measurement noises in the VANET environment [18,80,81]. The noise types, noise scenarios, and their simulation have been reported in our previous publication [70]. The acquisition algorithm presented in Reference [70] has been used to acquire its own context information in each vehicle.

4.1.3. Simulation of Message Losses

Because the performance of the applications is impacted by the messages loss rate, nine communication scenarios were simulated, each with different traffic density so as to represent different message loss rates. For each communication scenario, sixteen different traffic datasets have been conducted, each of which contains realistic driving regimes. A total of 144 experiments have been conducted to evaluate the proposed model. The Matlab [78] network simulation platform was used to simulate the IEEE 802.11p/WAVE communication among vehicles [82]. In each experiment, the maximum communication range used was 1 km, broadcasting frequency 10 Hz. Each vehicle generates a new mobility message at the beginning of every control channel interval (CCHI) [82]. Vehicles use the scheme presented in Reference [31] for efficient broadcasting and effective collection of neighboring vehicles' context information. The message arrival of vehicles was modeled as a random variable with Poisson distribution to simulate the message loss under different communication scenarios [83–86].

4.1.4. Simulation of Misbehaving Vehicles

Due to the absence of ground truth-labeled dataset for evaluating the intrusion detection systems in the VANET, the common procedure is to simulate the attacker's actions. Therefore, two types of context-driven attacks were simulated, basic attack and sophisticated attack. The basic attacks are

those that directly modify the mobility information messages without manipulating the context. This type of attack includes, but not limited to, positioning noises, position jumping, message suppression, cheating with context information, sudden position jumping, and random jumping attack. [52,87]. On the other hand, with a sophisticated attack, attackers are aware of the context and, consequently, can perform incremental jumps to carry out attacks [25].

4.2. Performance Metrics

Five performance metrics were used to evaluate the proposed EHCA-MDS, namely, detection accuracy, false positive rate (FPR), detection rate (DR), precision, and F-measure. These metrics are commonly used by researchers to evaluate the misbehavior detection models in VANET [29]. The first performance metric is the detection accuracy, which is the percentage of correctly classified vehicles. The second metrics is the false positive rate (*FPR*), which is the proportion of genuine vehicles that were misclassified. The third performance measure is the detection rate (*DR*), also called the recall or sensitivity, which is the proportion of the attacks that are correctly detected. The fourth performance measure is the precision (also called positive predictive value), which is a measure of statistical bias, and measures the correctly classified genuine vehicles over the total vehicles that were predicted as genuine vehicles. Finally, the fifth measure is the F-measure, which is the harmonic mean of the recall and the precision [88].

5. Results

In this section, the performance of the proposed EHCA-MDS, HCA-MDS, and DCA-MDS, in terms of the aforementioned performance measures, is presented. In HCA-MDS, the classification outputs of all statistical classifiers have been aggregated using the logical OR operator to obtain the final decision (the pre-detection phase). That is, a misbehaving vehicle is identified if the output of any one of the independent statistical classifiers is positive. The context-aware entity-centric MDS model (DCA-MDS) is similar to the hybrid model with the behavioral classifiers removed. The Ensembl-based Hybrid Context-Aware MDS (EHCA-MDS), Hybrid Context-Aware MDS (HCA-MDS), and Data-Centric Context-Aware MDS (DCA-MDS) were compared with the Bissmeyer's model [20], called the Entity-Centric Trust-based MDS (ECT-MDS) model and Stübing's MDS model [58]. Stübing's MDS [58] has been used as a baseline for evaluating misbehavior detection by several studies [20,29]. The results of the simulations are presented in Table 5. For each value in Table 5, the results were averaged over 144 runs (16 datasets \times 9 communication scenarios) with respect to the performance measure.

Table 5. The effectiveness of the proposed models.

Model	Accuracy%	FPR%	DR%	Precision%	Recall%	F-Measure%
EHCA-MDS (Proposed)	97.01	1.19	90.45	95.32	90.45	92.82
HCA-MDS (Proposed)	93.51	4.45	86.11	85.00	86.11	84.44
DCA-MDS (Proposed)	90.98	2.33	66.18	89.19	66.18	75.05
Bissmeyers' ECT-MDS [20]	74.79	2.98	30.65	83.50	30.65	44.49
Stübing's MDS [58]	87.37	4.79	62.55	86.91	62.55	71.60

The results in Table 5 show that EHCA-MDS has achieved the highest accuracy performance in terms of detection accuracy (see Figure 7a). That is, 97.01% of the vehicles were correctly classified by the EHCA-MDS compared to 93.01% for the HCA-MDS, 90.98% for the DCA-MDS, 74.79% for the ECT-MDS, and 87.37% for the baseline MDS model. In terms of the false positive rate, the EHCA-MDS achieved the lowest false alarms, 1.19% of benign vehicles were misclassified as attackers by the EHCA-MDS compared to 4.45%, 2.33%, 2.98%, and 4.79% of HCA-MDS, DCA-MDS, ECT-MDS, and MDS models, respectively. Furthermore, the ensemble-based approach, EHCA-MDS, achieved the lowest false positive rate (1.19%) compared to the independent-classifiers HCA-MDS and DCA-MDS and non-context-aware model ECT-MDS and Stübing's MDS (see Figure 7b), which achieved 4.45%, 2.33%, 2.98%, and 4.79%, respectively. Regarding the detection rate, an average of 90.45% of misbehaving

vehicles has been correctly identified by EHCA-MDS compared to 86.11%, 66.18%, 30.65%, and 62.55% for the HCA-MDS, CA-MDS, ECT-MDS, and MDS models, respectively. That is, EHCA-MDS has achieved the highest detection rate compared to the related work (see Figure 7c). Overall, EHCA-MDS has achieved the highest detection performance, which is 92.82% compared to 84.44%, 75.05%, 44.49%, and 71.60% for HCA-MDS, CA-MDS, ECT-MDS, and MDS models, respectively. The results in terms of F-measure shows that the proposed model achieves the highest trade-off between precision and recall (see Figure 7d). Table 6 summarized the improvement gained by the proposed EHCA-MDS compared to the other studied models HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline. In terms of the overall accuracy, the EHCA-MDS achieved 3.5, 6.03, 22.22, and 9.64 higher than HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline, respectively. That is, the accuracy has been improved by 3.74%, 36.67%, 29.71%, and 11.03%, as compared to HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline, respectively. The false-positive rate has been reduced to 1.19%, which becomes 73%, 36.67%, 60.07%, and 75.16% lower than that of HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline, respectively. The detection rate (DR) was improved by 5%, 36.67%, 195.11%, and 44.60% compared to HCA-MDS. In terms of precision, the proposed EHCA-MDS has been improved by 11.8% compared to HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline, respectively. The overall performance has been improved by 9.9% compared to HCA-MDS.

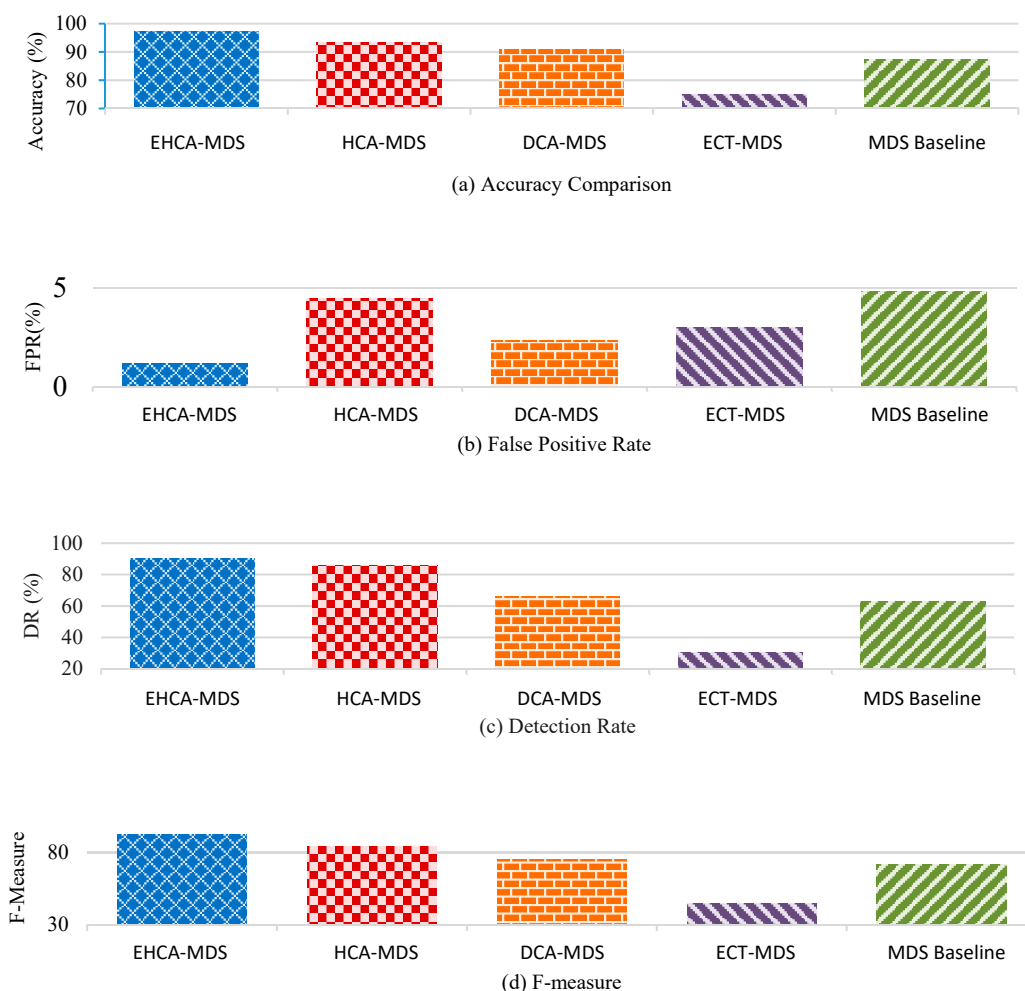


Figure 7. Average effectiveness of EHCA-MDS, HCA-MDS, CA-MDS, ECT-MDS and MDS baseline models in terms of the: (a) Accuracy, (b) False Positive Rate (FPR), (c) Detection Rate (DR), and (d) F-Measure.

Table 6. Improvement gained by the proposed EHCA-MDS model.

Model	Accuracy	FPR	DR	Precision	Recall	F-Measure
HCA-MDS (Proposed)	3.5 (3.74%)	−3.26 (73.26%)	4.34 (5.04%)	10.32 (12.14%)	4.34 (5.04%)	8.38 (9.92%)
DCA-MDS (Proposed)	6.03 (36.67%)	−1.14 (36.67%)	24.27 (36.67%)	6.13 (36.67%)	24.27 (36.67%)	17.77 (36.67%)
Bissmeyers' ETC-MDS [20]	22.22 (29.71%)	−1.79 (60.07%)	59.80 (195.11%)	11.82 (14.16%)	59.80 (195.11%)	48.33 (108.63%)
Stübing's MDS baseline [58]	9.64 (11.03%)	−3.60 (75.16%)	27.90 (44.60%)	8.41 (9.68%)	27.90 (44.60%)	21.22 (29.64%)

Figure 8 shows the performance comparison between the proposed EHCA-MDS, the independent classifiers-based approach (HCA-MDS), the data-centric context-aware-based approach (DCA-MDS), and the related non-hybrid non-context-aware models (ETC-MDS and MDS) under different context scenarios where the communication reliability degraded from ideal in the first scenario to the worst in the last scenario. The X-axis contains nine context scenarios. In each scenario, different messages' arrival rates that represent different communication statuses were used. The Y-axis contains the average results of replaying the 16 datasets listed in Table 4 under each context scenario in terms of the performance measures, namely, detection accuracy (Figure 8a), FPR (Figure 8b), DR (Figure 8c), and the F-Measure (Figure 8d).

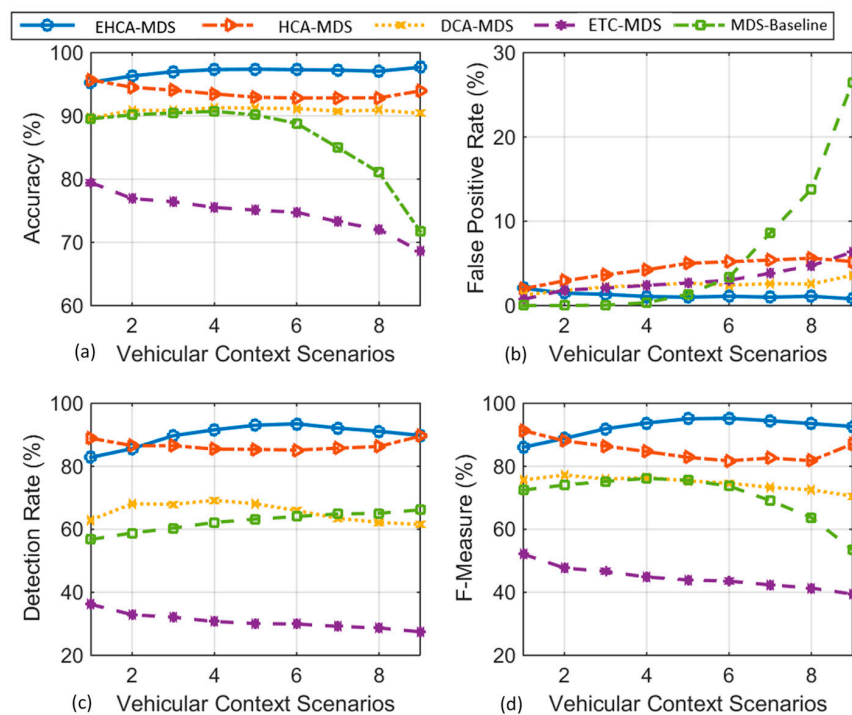
**Figure 8.** Detailed comparison of EHCA-MDS in terms of the: (a) Accuracy, (b) False Positive Rate (FPR), (c) Detection Rate (DR), and (d) F-Measure.

Figure 8a–d compare the impact of different vehicular contexts on the detection performance of each model. All the context-aware models (EHCA-MDS, HCA-MDS, and DCA-MDS) manage to keep the false positive rate low. Figure 8a shows the comparison of the accuracy performance of EHCA-MDS with the other implemented MDS models. Figure 8b compares the performance of EHCA-MDS in terms of the false-positive rate. EHCA-MDS has the lowest false positive rate in most scenarios, with a slight increase when communication becomes more unreliable. HCA-MDS has the highest false-positive rate among the context-aware models. Figure 8c compares the performance of EHCA-MDS with other MDS

models in terms of the detection rate. EHCA-MDS has the highest detection rate in most scenarios. In the first scenario where the communication was assumed ideal, the detection rate of the EHCA-MDS is lower than that of HCA-MDS. Figure 8d summarizes the overall performance of EHCA-EC-MDS compared to the other MDS models in terms of the F-measure. Particularly, EHCA-EC-MDS has the highest F-measure values in most scenarios. Only under the assumption of ideal communication, the overall performance of the EHCA-EC-MDS, in terms of F-measure, is lower than that of HCA-EC-MDS. However, the difference is insignificant.

6. Discussion

In this study, the ensemble hybrid context-aware misbehavior detection system model (EHCA-MDS) was introduced to address the adaptability issue of the MDS model with the highly dynamic vehicular context as well as the issue of insufficient context representation. To address the adaptability issue, Kalman and Hampel filters were used to build a hybrid multifaceted dynamic context reference utilizing neighboring vehicles' context information. Accordingly, hybrid multifaceted statistical models were constructed online and updated in a timely manner. In doing so, the misbehaving vehicles can be effectively detected locally and autonomously. That is, vehicles that deviated much from the context reference are considered suspicious. To address the issue of insufficient context representation, the output of the hybrid and multifaceted classifiers were used as new features to train an ensemble of machine-learning-based classifiers using the ensemble learning method. The random forest algorithm was used as the base classifier. The results from both statistical- and ensemble-based classifiers were combined using a weighted average function for more effective results. The results show the effectiveness and robustness of the proposed model for dynamic vehicular context. In Table 4, Table 5, Figure 7, and Figure 8, significant improvements were observed by the proposed EHCA-MDS model compared to the existing MDS models in all studied scenarios. The results confirm that the proposed ensemble-based model (EHCA-MDS) outperforms the existing models, including the independent hybrid model (HCA-MDS), the non-hybrid context-aware model (DCA-MDS), and the non-context-aware-based models (ECT-MDS and MDS) with high significance (see Table 4 and Figure 7).

The results in Table 4, Table 5, Figure 7, and Figure 8 indicate that the combination among several non-parametric, unsupervised-based online statistical classifiers with a supervised-based offline classifier consolidates the robustness and the effectiveness of VANET critical applications. The improvement achieved by the proposed model suggests that the consideration of the correlation among the hybrid and multifaceted features can capture the unseen behavior of the misbehaving vehicles. That is, sophisticated attacks that are aware of vehicular context and resemble benign vehicles can be easily identified when the correlations among features are considered. Considering the correlations between the outputs of individual classifiers provided effective context representation and worked as a pre-detection step that helps the ensemble-based classifiers to accurately distinguish the misbehaving vehicles from other benign ones. The lack of consideration of the correlation among the outputs of the statistical classifiers led to lower accuracy (see Figure 8a,c). This is because of the sophisticated attackers that are aware of the parameters of the statistical classifiers. Thus, those parameters can be used as attack thresholds to normalize attackers' behavior and avoid detection. To increase the detectability of the sophisticated attackers, the limits of the statistical classifiers can be shrunk. The problem, however, is that such shrinking leads to an increase in the false-positive rate. This interprets the high positive-rate resulted from the HCA-MDS (see Figure 8b) and led to reducing the overall performance (see Figure 8d). Both the non-context-aware and non-hybrid approaches (ECT-MDS [20] and MDS baseline [58]) are ineffective even for simple attack types (See Figure 8) due to the ignorance of the context and lack of considering the correlation among different features.

As shown in Figure 8a, EHCA-MDS remains stable at the highest accuracy among all the models in most of the scenarios. In the first scenario where the communication was assumed ideal, the accuracy of the HCA-MDS is slightly higher than the accuracy of the EHCA-MDS. However, the

EHCA-MDS is more robust than HCA-MDS when the communication becomes unreliable. This is due to the consideration of the relationship among the input features in the EHCA-MDS compared to the other models that treat the features separately, which has led to insufficient representation and thus lower accuracy. Although both DCA-MDS and HCA-MDS have a high false-positive rate compared to the EHCA-MDS, all context-aware models are relatively stable and robust to the context change (see Figure 8b). As shown in Figure 8b, EHCA-MDS is more stable and robust in terms of the false-positive rate than the other tested models. When the data is accurate, the baseline model has a relatively low false-positive rate compared to HCA-MDS and DCA-MDS. However, when the context becomes unreliable, or data accuracy is highly uncertain, the false-positive rate of the baseline MDS model increases rapidly.

From Figure 8c, it can be noted that in the first scenario where the communication was assumed ideal, the accuracy of the HCA-MDS is slightly higher than the accuracy of the EHCA-MDS. This is because the ensemble-based learning-based classifiers could not well distinguish between the patterns of the sophisticated attack due to its high similarity with the behavior of benign vehicles in the ideal context scenarios. This explains the drop in the accuracy under the first scenario, as shown in Figure 8a. The detection rate remains stable as the communication becomes highly unreliable as their detection rate stays stable under different context scenarios (see scenarios 6, 7, 8, and 9 in Figure 8c). It can be observed that neither context-aware nor non-context-aware was able to detect sophisticated attacks effectively. The non-context-aware schemes fail to achieve a balance between the detection rate and the false positive rate. That is, both the false positive rate and the detection rate of the non-context-aware (Bissmeyers' ECT-MDS and Stübing's MDS) approach increases dramatically when the communication becomes more unreliable. In contrast, the context-aware approach (EHCA-MD, HCA-MDS, and DCA-MDS) has a better trade-off between the detection rate and the false positive rate (see Figure 8b,c). It can be observed that the EHCA-MDS is more effective in terms of the trade-off between detection rate and false-positive rate among all studied models. Compared with the non-hybrid approaches, EHCA-MDS and HCA-MDS are more effective and stable when making a trade-off between the false positive rate and the detection rate compared with the data-centric and non-context-aware models (DCA-MDS).

From Figure 8d, one can observe that the overall performance in terms of F-measure shows that EHCA-EC-MDS is more robust and effective than the other studied models. Although the overall performance of the EHCA-EC-MDS is lower than that of HCA-EC-MDS under the assumption of ideal communication, the difference is insignificant. This is due to the drop in the detection rate, as shown in Figure 8c. The drop in the detection rate is due to the high similarity between attack behavior and normal vehicle behavior. The weights that were given to the classifiers during the decision-making process worked as a penalty in high message arrival rate scenarios. This interprets the improvement in the detection rate when the message arrival rate decreases. Although the difference is insignificant, one idea to solve such a problem is by adapting the classifiers' weights according to the context. This issue has been left for future work.

To sum up, the performance achieved by the ensemble-based learning approach, i.e., EHCA-MDS, was better than that of the independent statistical classifiers-based approach (represented by HCA-MDS, DCA-MDS, ECT-MDS, and MDS baseline) due to its consideration of the correlation among the outputs of the statistical classifiers. Meanwhile, the performance achieved by the hybrid-based approach, i.e., HCA-MDS, outperformed the non-hybrid approach (DCA-MDS, Bissmeyers' ECT-MDS, and Stübing's MDS) due to involving the multifaceted classifiers, which addressed the direct relationship between the behavioral- and data-centric-based features. For example, the attacker who tries to share false events will be keen to increase its broadcasting rate to get more chance over the congested communication channel. On average, HCA-MDS has improved the overall performance by 17.93% compared to Stübing's MDS model. The proposed context-aware models EHCA-MDS, HCA-MDS, and DCA-MDS can adapt to dynamic vehicular context achieving higher performance than that of the none-context aware models (ECT-MDS and MDS baseline). Table 7 summarizes the important features that have led

to the performance improvement in terms of F-measure achieved by the proposed EHCA-MDS model compared to the existing models.

Table 7. Features comparison among the studied and proposed models.

Model	Ensemble	Hybrid *	Context-Aware	Data-Centric	Trust-Based	Performance
EHCA-MDS	✓	✓	✓	✓		92.82
HCA-MDS		✓	✓	✓		84.44
DCA-MDS		✓	✓	✓		75.05
ECT-MDS		✓		✓	✓	71.60
MDS Baseline				✓		44.49

* Hybrid \equiv Data_Centric + Behavioural, Performance = F-measure.

7. Conclusions and Future Work

In this paper, an ensemble hybrid context-aware misbehavior detection model was developed. EHCA-MDS follows the concepts of the wisdom of crowds and the power of diversity, in which crowds of random and uncorrelated classifiers have been used with multifaceted independent classifiers. These classifiers work together to detect the different types of misbehaving vehicles that share false mobility messages. Both unsupervised and supervised approaches were used to construct the crowds of the classifiers. The robust statistical-based outlier classifiers were constructed and updated online to detect emerging attacks. Meanwhile, the random forest algorithm was used to train crowds of smart classifiers to predict the vehicles' class based on learning from the previously seen pattern. The model consists of hybrid and multifaceted statistical classifiers that work together to detect sophisticated attacks that the single-classifier approach cannot detect accurately. A multifaceted and hybrid context-reference model is constructed and updated in a timely manner by analyzing the spatial and temporal correlation among neighboring mobility information using Kalman and the Hampel filter-based algorithms. Significant improvement has been observed by the proposed model in terms of detecting misbehaving vehicles under different context scenarios. The EHCA-MDS model has manifested its adaptability and robustness even under unreliable communication and heterogeneous noise environment. Results show that they can make a trade-off between precision and recall effectively. On average, the overall performance of the proposed EHCA-MDS model has improved by 10% better than the hybrid model (HCA-MDS) and 37% better than the data-centric model (DCA-MDS).

In the future, the use of the supervised method can be replaced by unsupervised methods, due to the lack of labeled data and the dynamic nature of the normal context. Further research is required to investigate which unsupervised learning techniques are more effective and efficient for local and autonomous detection. In addition, the proposed EHCA-MDS relies on first-hand information (the mobility information shared by neighboring vehicles) to detect short-term misbehavior. However, second-hand information (the output of the misbehavior detection shared between neighboring vehicles) can be used to make long-term detection.

Author Contributions: Conceptualization, F.A.G. and A.Z.; Data curation, F.A.G.; Formal analysis, F.A.G.; Funding acquisition, M.A.M., A.Z., A.A. and W.B.; Investigation, F.A.G. and B.A.S.A.; Methodology, F.A.G., A.Z.; Project administration, M.A.M.; Resources, M.A.M., A.A., and W.B.; Software, F.A.G., and A.A.; Supervision, M.A.M. and A.Z.; Validation, F.A.G., B.A.S.A., and W.B.; Visualization, F.A.G., W.B., and A.Z.; Writing—original draft, F.A.G.; Writing—review & editing, F.A.G., B.A.S.A., W.B. and A.Z.

Funding: His research was funded by the Ministry of Higher Education (MOHE) and Research Management Centre (RMC) at the Universiti Teknologi Malaysia (UTM) under Post-Doctoral Fellowship Scheme (VOT Q.J130000.21A2.04E00).

Acknowledgments: Our deep gratitude is extended to the Ministry of Higher Education (MOHE), Malaysian International Scholarship (MIS), and Cybersecurity Research lab, School of Computing at the Universiti Teknologi Malaysia (UTM) for their unlimited support throughout this study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. WHO. 10 Facts on Global Road Safety. 2013. Available online: <http://www.who.int/features/factfiles/roadsafety/en/> (accessed on 20 June 2019).
2. Wahab, O.A.; Mourad, A.; Otrók, H.; Bentahar, J. CEAP: SVM-based intelligent detection model for clustered vehicular ad hoc networks. *Expert Syst. Appl.* **2016**, *50*, 40–54. [\[CrossRef\]](#)
3. Arnott, R.; de Palma, A.; Lindsey, R. Economics of a bottleneck. *J. Urban Econ.* **1990**, *27*, 111–130. [\[CrossRef\]](#)
4. Sweet, M. Does traffic congestion slow the economy? *J. Plan. Lit.* **2011**, *26*, 391–404. [\[CrossRef\]](#)
5. Williams, B.M.; Guin, A. Traffic Management Center Use of Incident Detection Algorithms: Findings of a Nationwide Survey. *IEEE Trans. Intell. Transp. Syst.* **2007**, *8*, 351–358. [\[CrossRef\]](#)
6. Vahdat-Nejad, H.; Ramazani, A.; Mohammadi, T.; Mansoor, W. A survey on context-aware vehicular network applications. *Veh. Commun.* **2016**, *3*, 43–57. [\[CrossRef\]](#)
7. Heijden, R.W.; Kargl, F. *Open Issues in Differentiating Misbehavior and Anomalies for VANETs*; Vehicular Lab, University of Luxembourg: Luxembourg, 2014.
8. Santamaria, A.F.; Sottile, C.; De Rango, F.; Voznak, M. Road Safety Alerting System with Radar and GPS Cooperation in a VANET Environment. In Proceedings of the Wireless Sensing, Localization, and Processing IX, Baltimore, MD, USA, 5–9 May 2014; Volume 9103.
9. Uzcategui, R.; Acosta-Marum, G. Wave: A tutorial. *IEEE Commun. Mag.* **2009**, *47*, 126–133. [\[CrossRef\]](#)
10. Dietzel, S. Graph-Based Metrics for Insider Attack Detection in VANET Multihop Data Dissemination Protocols. *IEEE Trans. Veh. Technol.* **2013**, *62*, 1505–1518. [\[CrossRef\]](#)
11. Huang, C.M.; Lin, S.Y. Cooperative vehicle collision warning system using the vector-based approach with dedicated short range communication data transmission. *IET Intell. Transp. Syst.* **2014**, *8*, 124–134. [\[CrossRef\]](#)
12. Le, L.; Festag, A.; Baldessari, R.; Zhang, W. Vehicular wireless short-range communication for improving intersection safety. *IEEE Commun. Mag.* **2009**, *47*, 104–110.
13. Kato, S.; Tsugawa, S.; Tokuda, K.; Matsui, T.; Fujii, H. Vehicle control algorithms for cooperative driving with automated vehicles and intervehicle communications. *IEEE Trans. Intell. Transp. Syst.* **2002**, *3*, 155–161. [\[CrossRef\]](#)
14. Milan, V.; Shladover, S.E.; Spring, J.; Nowakowski, C.; Kawazoe, H.; Nakamura, M. Cooperative Adaptive Cruise Control in Real Traffic Situations. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 296–305. [\[CrossRef\]](#)
15. Li, L.; Wen, D.; Zheng, N.N.; Shen, L.C. Cognitive Cars: A New Frontier for ADAS Research. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 395–407. [\[CrossRef\]](#)
16. NHTSA. *Vehicle Safety Communications Project Task 3 Final Report, Identify Intelligent Vehicle Safety Applications Enabled by Dsrc*; US Department of Transportation: Washington, DC, USA, 2005.
17. Stübting, H. Car-to-X Communication: System Architecture and Applications. In *Multilayered Security and Privacy Protection in Car-to-X Networks*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2013; pp. 9–19.
18. Liu, K.; Lim, H.B.; Frazzoli, E.; Ji, H.; Lee, V.C.S. Improving positioning accuracy using GPS pseudorange measurements for cooperative vehicular localization. *IEEE Trans. Veh. Technol.* **2014**, *63*, 2544–2556. [\[CrossRef\]](#)
19. Sakiz, F.; Sen, S. A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV. *Ad Hoc Netw.* **2017**, *61*, 33–50. [\[CrossRef\]](#)
20. Bissmeyer, N.; Michael, W.; Frank, K. Misbehavior Detection and Attacker Identification in Vehicular Ad-Hoc Networks. Ph.D. Thesis, Technische Universität Darmstadt, Darmstadt, Germany, 2014.
21. Leinmuller, T.; Schoch, E.; Kargl, F.; Maihöfer, C. Influence of falsified position data on geographic ad-hoc routing. In *Security and Privacy in Ad-hoc and Sensor Networks*; Springer: Berlin, Germany, 2005; pp. 102–112.
22. Leinmuller, T.; Schoch, E. Greedy routing in highway scenarios: The impact of position faking nodes. In Proceedings of the Workshop On Intelligent Transportation (WIT 2006), Hamburg, Germany, 14–15 March 2006.
23. Grover, J.; Gaur, M.S.; Laxmi, V. Position Forging Attacks in Vehicular Ad Hoc Networks: Implementation, Impact and Detection. In Proceedings of the 7th International Wireless Communications and Mobile Computing Conference (IWCMC), Istanbul, Turkey, 4–8 July 2011.
24. Blum, J.; Eskandarian, A. The threat of intelligent collisions. *IT Prof.* **2004**, *6*, 24–29. [\[CrossRef\]](#)
25. Lo, N.-W.; Tsai, H.-C. Illusion Attack on VANET Applications—A Message Plausibility Problem. In Proceedings of the IEEE Globecom Workshops, Washington, DC, USA, 26–30 November 2007.

26. Yu, B.; Xu, C.Z.; Xiao, B. Detecting Sybil attacks in VANETs. *J. Parallel Distrib. Comput.* **2013**, *73*, 746–756. [\[CrossRef\]](#)
27. Kakkasageri, M.S.; Manvi, S.S. Information management in vehicular ad hoc networks: A review. *J. Netw. Comput. Appl.* **2014**, *39*, 334–350. [\[CrossRef\]](#)
28. Chen, S.; Wang, W.; van Zuylen, H. A comparison of outlier detection algorithms for ITS data. *Expert Syst. Appl.* **2010**, *37*, 1169–1178. [\[CrossRef\]](#)
29. Firl, J.; Stübing, H.; Huss, S.A.; Stiller, C. MARV-X: Applying Maneuver Assessment for Reliable Verification of Car-to-X Mobility Data. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1301–1312. [\[CrossRef\]](#)
30. Van der Heijden, R.W.; Stefan, D.; Tim, L.; Frank, K. Survey on Misbehavior Detection in Cooperative Intelligent Transportation Systems. *IEEE Commun. Surv. Tutor.* **2019**, *21*, 779–811. [\[CrossRef\]](#)
31. Ghaleb, F.A.; Zainal, A.; Rassam, M.A.; Saeed, F. Driving-situation-aware adaptive broadcasting rate scheme for vehicular ad hoc network. *J. Intell. Fuzzy Syst.* **2018**, *35*, 423–438. [\[CrossRef\]](#)
32. Zhengming, L.; Chunxiao, C.; Wong, D. AWF-NA: A Complete Solution for Tampered Packet Detection in VANETs. In Proceedings of the IEEE GLOBECOM 2008—2008 Global Telecommunications Conference, New Orleans, LO, USA, 30 November–4 December 2008.
33. Hortelano, J.; Ruiz, J.C.; Manzoni, P. Evaluating the Usefulness of Watchdogs for Intrusion Detection in VANETs. In Proceedings of the 2010 IEEE International Conference on Communications Workshops, Capetown, South Africa, 23–27 May 2010.
34. Daeinabi, A.; Rahbar, A.G. Detection of malicious vehicles (DMV) through monitoring in Vehicular Ad-Hoc Networks. *Multimed. Tools Appl.* **2013**, *66*, 325–338. [\[CrossRef\]](#)
35. Wahab, O.A.; Otok, H.; Mourad, A. A cooperative watchdog model based on Dempster-Shafer for detecting misbehaving vehicles. *Comput. Commun.* **2014**, *41*, 43–54. [\[CrossRef\]](#)
36. Zhang, Y.; Lazos, L.; Kozma, W. AMD: Audit-Based Misbehavior Detection in Wireless Ad Hoc Networks. *IEEE Trans. Mob. Comput.* **2016**, *15*, 1893–1907. [\[CrossRef\]](#)
37. Ho, Y.-H.; Lin, C.-H.; Chen, L.-J. On-demand Misbehavior Detection for Vehicular Ad Hoc Network. *Int. J. Distrib. Sens. Netw.* **2016**, *12*, 1550147716673928. [\[CrossRef\]](#)
38. Ghosh, M.; Varghese, A.; Gupta, A.; Kherani, A.A.; Muthaiah, S.N. Detecting misbehaviors in VANET with integrated root-cause analysis. *Ad Hoc Netw.* **2010**, *8*, 778–790. [\[CrossRef\]](#)
39. Grover, J.; Laxmi, V.; Gaur, M.S. Misbehavior Detection Based on Ensemble Learning in VANET. In *International Conference on Advanced Computing, Networking and Security. ADCONS 2011, Surathkal, India, December 16–18, 2011, Revised Selected Papers*; Thilagam, P.S., Pais, A.R., Chandrasekaran, K., Balakrishnan, N., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 602–611.
40. Hussain, R.; Kim, S.; Oh, H. Privacy-Aware VANET Security: Putting Data-Centric Misbehavior and Sybil Attack Detection Schemes into Practice. In *Information Security Applications*; Lee, D., Yung, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 296–311.
41. Zaidi, K.; Milojevic, M.; Rakocevic, V.; Rajarajan, M. Data-centric Rogue Node Detection in VANETs. In Proceedings of the IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications, Beijing, China, 24–26 September 2014.
42. Ruj, S.; Cavenaghi, M.A.; Huang, Z.; Nayak, A.; Stojmenovic, I. On data-centric misbehavior detection in VANETs. In Proceedings of the IEEE Vehicular Technology Conference (VTC Fall), San Francisco, CA, USA, 5–8 September 2011.
43. Raya, M.; Papadimitratos, P.; Gligor, V.D.; Hubaux, J.-P. On Data-Centric Trust Establishment in Ephemeral Ad Hoc Networks. In Proceedings of the IEEE INFOCOM 2008—The 27th Conference on Computer Communications, Phoenix, AZ, USA, 13–18 April 2008.
44. Zaidi, K.; Milojevic, M.B.; Rakocevic, V.; Nallanathan, A.; Rajarajan, M. Host-Based Intrusion Detection for VANETs: A Statistical Approach to Rogue Node Detection. *IEEE Trans. Veh. Technol.* **2016**, *65*, 6703–6714. [\[CrossRef\]](#)
45. Grover, J.; Kumar, D.; Sargurunathan, M.; Gaur, M.S.; Laxmi, V. Performance Evaluation and Detection of Sybil Attacks in Vehicular Ad-Hoc Networks. *Recent Trends Netw. Secur. Appl.* **2010**, *89*, 473–482.
46. Dietzel, S.; van der Heijden, R.; Decke, H.; Kargl, F. A flexible, subjective logic-based framework for misbehavior detection in V2V networks. In Proceedings of the IEEE 15th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), Sydney, Australia, 19 June 2014.

47. Golle, P.; Greene, D.; Staddon, J. Detecting and correcting malicious data in VANETs. In Proceedings of the 1st ACM International Workshop on Vehicular ad hoc Networks, Philadelphia, PA, USA, 1 October 2004; pp. 29–37.
48. Yan, G.J.; Olariu, S.; Weigle, M.C. Providing Location Security in Vehicular Ad Hoc Networks. *IEEE Wirel. Commun.* **2009**, *16*, 48–53. [\[CrossRef\]](#)
49. Bissmeyer, N.; Njeukam, J.; Petit, J.; Bayarou, K.M. Central misbehavior evaluation for VANETs based on mobility data plausibility. In Proceedings of the Ninth ACM International Workshop on Vehicular Inter-networking, Systems, and Applications, Low Wood Bay, Lake District, UK, 25 June 2012; pp. 73–82.
50. Leinmuller, T.; Schoch, E.; Kargl, F.; Maihöfer, C. Decentralized position verification in geographic ad hoc routing. *Secur. Commun. Netw.* **2010**, *3*, 289–302. [\[CrossRef\]](#)
51. Bissmeyer, N.; Stresing, C.; Bayarou, K.M. Intrusion detection in VANETs through verification of vehicle movement data. In Proceedings of the IEEE Vehicular Networking Conference (VNC), Jersey City, NJ, USA, 13–15 December 2010.
52. Bissmeyer, N.; Schröder, K.H.; Petit, J.; Mauthofer, S.; Bayarou, K.M. Short paper: Experimental analysis of misbehavior detection and prevention in VANETs. In Proceedings of the Fifth IEEE Vehicular Networking Conference (VNC), Boston, MA, USA, 16–18 December 2013; pp. 198–201.
53. Palomar, E.; de Fuentes, J.M.; González-Tablas, A.I.; Alcaide, A. Hindering false event dissemination in VANETs with proof-of-work mechanisms. *Transp. Res. Part C Emerg. Technol.* **2012**, *23*, 85–97. [\[CrossRef\]](#)
54. Van der Heijden, R.W.; Dietzel, S.; Leinmüller, T.; Kargl, F. Survey on Misbehavior Detection in Cooperative Intelligent Transportation Systems. *arXiv* **2016**, arXiv:1610.06810. [\[CrossRef\]](#)
55. Schmidt, R.K.; Leinmüller, T.; Schoch, E.; Held, A.; Schäfer, G. Vehicle behavior analysis to enhance security in vanets. In Proceedings of the 4th IEEE Vehicle-to-Vehicle Communications Workshop (V2VCOM2008), Eindhoven, The Netherlands, 3 June 2008.
56. Leinmuller, T.; Schoch, E.; Kargl, F. Position verification approaches for vehicular ad hoc networks. *IEEE Wirel. Commun.* **2006**, *13*, 16–21. [\[CrossRef\]](#)
57. Jaeger, A.; Bißmeyer, N.; Stübing, H.; Huss, S.A. A Novel Framework for Efficient Mobility Data Verification in Vehicular Ad-hoc Networks. *Int. J. Intell. Transp. Syst. Res.* **2012**, *10*, 11–21. [\[CrossRef\]](#)
58. Stübing, H.; Jaeger, A.; Schmidt, C.; Huss, S.A. Verifying mobility data under privacy considerations in car-to-x communication. In Proceedings of the 17th ITS World Congress, Busan, Korea, 25–29 October 2010.
59. Stübing, H.; Firl, J.; Huss, S.A. A two-stage verification process for Car-to-X mobility data based on path prediction and probabilistic maneuver recognition. In Proceedings of the 2011 IEEE Vehicular Networking Conference (VNC), Amsterdam, The Netherlands, 14–16 November 2011.
60. Ghaleb, F.A.; Razzaque, M.A.; Isnin, I.F. Security and privacy enhancement in VANETs using mobility pattern. In Proceedings of the 2013 Fifth International Conference on Ubiquitous and Future Networks (ICUFN), Da Nang, Vietnam, 2–5 July 2013.
61. Firl, J.; Stübing, H.; Huss, S.A.; Stiller, C. Predictive maneuver evaluation for enhancement of Car-to-X mobility data. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Alcalá de Henares, Spain, 3–7 June 2012.
62. Ghaleb, F.A.; Zainal, A.; Rassam, M.A. Mobility information estimation algorithm using Kalman-filter for vehicular ad hoc networks. *Int. J. Inf. Comput. Secur.* **2016**, *8*, 221–240. [\[CrossRef\]](#)
63. Stübing, H. Facility Layer Security: Mobility Data Verification. In *Multilayered Security and Privacy Protection in Car-to-X Networks*; Springer Fachmedien Wiesbaden: Berlin, Germany, 2013; pp. 45–80.
64. Jwo, D.-J.; Cho, T.-S. A practical note on evaluating Kalman filter performance optimality and degradation. *Appl. Math. Comput.* **2007**, *193*, 482–505. [\[CrossRef\]](#)
65. Punzo, V.; Borzacchiello, M.T.; Ciuffo, B. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transp. Res. Part C Emerg. Technol.* **2011**, *19*, 1243–1262. [\[CrossRef\]](#)
66. Huang, H.; Zhang, D.; Zhu, Y.; Li, M.; Wu, M.-Y. A Metropolitan Taxi Mobility Model from Real GPS Traces. *J. Univers. Comput. Sci.* **2012**, *18*, 1072–1092.
67. Ghaleb, F.A.; Maarof, M.A.; Zainal, A.; Al-Rimy, B.A.S.; Saeed, F.; Al-Hadhrani, T. Hybrid and Multifaceted Context-Aware Misbehavior Detection Model for Vehicular Ad Hoc Network. *IEEE Access* **2019**, *7*, 159119–159140. [\[CrossRef\]](#)

68. Ghaleb, F.A.; Zainal, A.; Rassam, M.A.; Mohammed, F. An effective misbehavior detection model using artificial neural network for vehicular ad hoc network applications. In Proceedings of the 2017 IEEE Conference on Application, Information and Network Security (AINS), Miri, Malaysia, 13–14 November 2017.
69. Ghaleb, F.A.; Maarof, M.A.; Zainal, A.; Rassam, M.A.; Saeed, F.; Alsaedi, M. Context-Aware Data-Centric Misbehaviour Detection Scheme for Vehicular Ad Hoc Networks using Sequential Analysis of the Temporal and Spatial Correlation of the Consistency between the Cooperative Awareness Messages. *Veh. Commun.* **2019**, *20*, 100186. [\[CrossRef\]](#)
70. Ghaleb, F.A.; Zainal, A.; Rassam, M.A.; Abraham, A. Improved vehicle positioning algorithm using enhanced innovation-based adaptive Kalman filter. *Pervasive Mob. Comput.* **2017**, *40*, 139–155. [\[CrossRef\]](#)
71. Engoulou, R.G.; Bellaïche, M.; Pierre, S.; Quintero, A. VANET security surveys. *Comput. Commun.* **2014**, *44*, 1–13. [\[CrossRef\]](#)
72. Boukerche, A.; Oliveira, H.A.B.F.; Nakamura, E.F.; Loureiro, A.A.F. Localization systems for wireless sensor networks. *IEEE Wirel. Commun.* **2007**, *14*, 6–12. [\[CrossRef\]](#)
73. Krell, M.M.; Tabie, M.; Woehrl, H.; Kirchner, E.A. Memory and Processing Efficient Formula for Moving Variance Calculation in EEG and EMG Signal Processing. In Proceedings of the International Congress on Neurotechnology, Electronics and Informatics (NEUROTECHNIX-2013), Vilamoura, Portugal, 19–20 September 2013.
74. FHWA. Next Generation Simulation (NGSIM) Vehicle Trajectories Dataset. 2006. Available online: <http://ngsim-community.org/> (accessed on 17 December 2015).
75. Hou, Y.; Edara, P.; Sun, C. Modeling Mandatory Lane Changing Using Bayes Classifier and Decision Trees. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 647–655. [\[CrossRef\]](#)
76. Tian, X.Y.; Liu, Y.H.; Wang, J.; Deng, W.W.; Oh, H. Computational Security for Context-Awareness in Vehicular Ad-Hoc Networks. *IEEE Access* **2016**, *4*, 5268–5279. [\[CrossRef\]](#)
77. Erskine, S.K.; Elleithy, K.M. Real-Time Detection of DoS Attacks in IEEE 802.11 p Using Fog Computing for a Secure Intelligent Vehicular Network. *Electronics* **2019**, *8*, 776. [\[CrossRef\]](#)
78. Weideman, J.A.; Reddy, S.C. A MATLAB differentiation matrix suite. *ACM Trans. Math. Softw.* **2000**, *26*, 465–519. [\[CrossRef\]](#)
79. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [\[CrossRef\]](#)
80. Drawil, N.M.; Basir, O. Intervehicle-Communication-Assisted Localization. *IEEE Trans. Intell. Transp. Syst.* **2010**, *11*, 678–691. [\[CrossRef\]](#)
81. Parker, R.; Valaee, S. Vehicular Node Localization Using Received-Signal-Strength Indicator. *IEEE Trans. Veh. Technol.* **2007**, *56*, 3371–3380. [\[CrossRef\]](#)
82. Qiu, H.J.F.; Ho, I.W.-H.; Tse, C.K.; Xie, Y. A Methodology for Studying 802.11p VANET Broadcasting Performance with Practical Vehicle Distribution. *IEEE Trans. Veh. Technol.* **2015**, *64*, 4756–4769. [\[CrossRef\]](#)
83. Mcquighan, P. *SIMULATING THE POISSON PROCESS*; Department of Mathematics, University of Chicago: Chicago, IL, USA, 2010.
84. Park, Y.; Kim, H. Application-Level Frequency Control of Periodic Safety Messages in the IEEE WAVE. *IEEE Trans. Veh. Technol.* **2012**, *61*, 1854–1862. [\[CrossRef\]](#)
85. Ghafoor, K.; Lloret, J.; Bakar, K.A.; Sadiq, A.S.; Mussa, S.A.B. Beaconing Approaches in Vehicular Ad Hoc Networks: A Survey. *Wirel. Pers. Commun.* **2013**, *73*, 885–912. [\[CrossRef\]](#)
86. Ma, X.; Zhang, J.; Yin, X.; Trivedi, K.S. Design and Analysis of a Robust Broadcast Scheme for VANET Safety-Related Services. *IEEE Trans. Veh. Technol.* **2012**, *61*, 46–61. [\[CrossRef\]](#)
87. Nikaein, N.; Kanti, S.D.; Marecar, I.; Bonnet, C. Application Distribution Model and Related Security Attacks in VANET. In Proceedings of the International Conference on Graphic and Image Processing (Icgip 2012), Singapore, 5–7 October 2013.
88. Chen, Y.-M.; Wei, Y.-C. A beacon-based trust management system for enhancing user centric location privacy in VANETs. *J. Commun. Netw.* **2013**, *15*, 153–163. [\[CrossRef\]](#)

