Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.040784

ARTICLE

Push-Based Content Dissemination and Machine Learning-Oriented Illusion Attack Detection in Vehicular Named Data Networking

Arif Hussain Magsi[1], Ghulam Muhammad[2,*], Sajida Karim[3], Saifullah Memon[1] and Zulfiqar Ali[4]

[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Post and Telecommunication, Beijing, 100876, China

[2]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, 11543, Saudi Arabia

[3]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, 150001, China

[4]School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK

*Corresponding Author: Ghulam Muhammad. Email: ghulam@ksu.edu.sa

Received: 03 March 2023 Accepted: 21 June 2023 Published: 08 October 2023

ABSTRACT

Recent advancements in the Vehicular Ad-hoc Network (VANET) have tremendously addressed road-related challenges. Specifically, Named Data Networking (NDN) in VANET has emerged as a vital technology due to its outstanding features. However, the NDN communication framework fails to address two important issues. The current NDN employs a pull-based content retrieval network, which is inefficient in disseminating crucial content in Vehicular Named Data Networking (VNDN). Additionally, VNDN is vulnerable to illusion attackers due to the administrative-less network of autonomous vehicles. Although various solutions have been proposed for detecting vehicles' behavior, they inadequately addressed the challenges specific to VNDN. To deal with these two issues, we propose a novel push-based crucial content dissemination scheme that extends the scope of VNDN from pull-based content retrieval to a push-based content forwarding mechanism. In addition, we exploit Machine Learning (ML) techniques within VNDN to detect the behavior of vehicles and classify them as attackers or legitimate. We trained and tested our system on the publicly accessible dataset Vehicular Reference Misbehavior (VeReMi). We employed five ML classification algorithms and constructed the best model for illusion attack detection. Our results indicate that Random Forest (RF) achieved excellent accuracy in detecting all illusion attack types in VeReMi, with an accuracy rate of 100% for type 1 and type 2, 96% for type 4 and type 16, and 95% for type 8. Thus, RF can effectively evaluate the behavior of vehicles and identify attacker vehicles with high accuracy. The ultimate goal of our research is to improve content exchange and secure VNDN from attackers. Thus, our ML-based attack detection and prevention mechanism ensures trustworthy content dissemination and prevents attacker vehicles from sharing misleading information in VNDN.

KEYWORDS

Named data networking; vehicular networks; pull-push; illusion attack; machine learning

1 Introduction

The continuous growth in the number of traditional vehicles has posed a significant threat to the safety of both drivers and passengers. According to an estimation in 2015, vehicles will increase twofold in the next 10 to 20 years [1]. The World Health Organization (WHO) predicts that road accidents will be the fifth leading cause of mortality by 2030 [2]. To cope with these issues, the Vehicular Ad-hoc Network (VANET) [3] has emerged as an indispensable solution for mitigating road accidents and traffic congestion. VANET-enabled vehicles can communicate as Vehicle-to-Vehicle (V2V), Vehicle-to-Pedestrians (V2P), and Vehicle-to-Infrastructure (V2I) [4] through Dedicated Short-Range Communication (DSRC), as illustrated in Fig. 1.

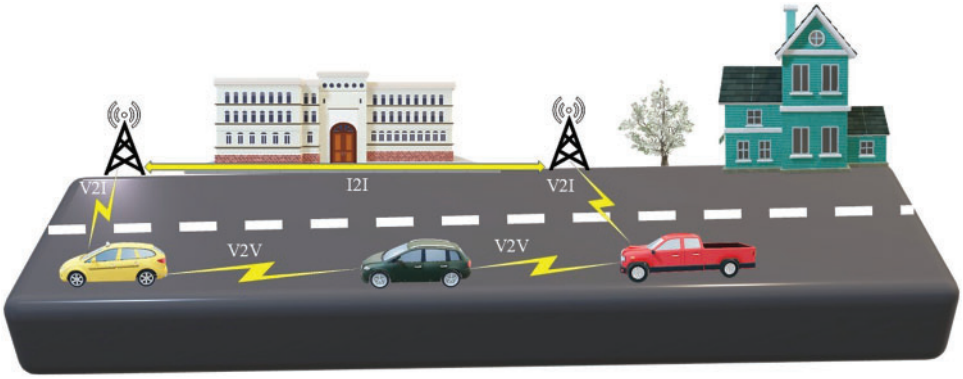

Figure 1: Content forwarding in VNDN

VANET-enabled vehicles are different from traditional vehicles; they are empowered with an Onboard Unit (OBU), which has computation, storage, and communication capabilities. However, the current VANET encounters several communication challenges due to the traditional Transmission Control Protocol/Internet Protocol (TCP/IP) [5]. TCP/IP is a host-centric, connection-oriented, and end-to-end communication system. TCP/IP must establish a connection before data transmission, which cannot support high mobility and intermittent connectivity in VANET [6]. As a result, Quality of Service (QoS) [7] leads to network degradation. The VANET is an extremely unbounded and intermittent network where nodes are interested in content rather than the source.

Alternatively, Named Data Networking (NDN) [8] has appeared as a special implementation of Information-Centric Networks (ICN) [9]. NDN is poised to be a promising candidate for next-generation internet architecture that considers content a "first-class citizen". This advancement has several benefits, including reduced overhead, lower latency, energy efficiency, and eliminating sender-receiver dependency. Particularly, the in-network content caching has significantly reduced latency in NDN. In contrast to traditional IP-based network architecture, NDN leverages content names to retrieve data without considering the content provider or its location. The naming structure in NDN follows a unique hierarchical structure separated by "/". In NDN, the content transmission uses two key packet types: *interest* and *data*.

NDN employs a pull-based request-response model to retrieve the content, where the content is delivered in response to specific consumer requests. Each vehicle in the Vehicular NDN (VNDN) holds three different data structures. **1. Content Store (CS):** The CS is an onboard storage component that temporarily caches the data to facilitate neighboring vehicles to access the content locally. The CS serves the consumer when the requested interest packet matches the content in the CS. **2. Pending Interest Table (PIT):** It is a table that maintains a record of all the interest packets and their associated

interfaces that have not yet been delivered to the consumers. **3. Forwarding Information Base (FIB):** It maintains the name prefixes of the next hop interface where interest is to be forwarded.

Despite its numerous capabilities, NDN has two significant limitations. The first limitation is associated with its reliance on a pull-based communication mechanism. In this approach, the consumer expresses interest in a specific piece of data by transmitting an interest packet. The network then forwards the interest packet to the producer, which responds to the consumer with a data packet. It is important to note that the producer can only respond to the content once the consumer initiates an interest request. However, this request-response-based communication mechanism fails to disseminate crucial content to the vehicles in an emergency and introduces an undesirable delay. In VNDN, a vehicle should be able to disseminate the crucial content among its neighboring vehicles without considering interest packet requests. On the other hand, push-based communication is a producer-driven approach where the producer initiates data transmission without waiting for a consumer's request.

Several studies in the literature have proposed push-based content dissemination approaches to enable the vehicles to propagate crucial data. The authors in [10] introduced a push-based layer-2 beacon forwarding mechanism among roadside units (RSUs) via one more intermediate vehicle. However, including payload in beacon messages within VNDN leads to increased latency. Similarly, reference [11] considered betweenness centrality to forward the push-based content using Publish-Subscribe (pub-sub) scheme. In pub-sub communication, consumers subscribe to the topic of their interest. When a producer publishes topic-related data, the network delivers it to all subscribed consumers. However, pub-sub cannot be suitable for disseminating crucial content in VNDN. In VNDN, every node must receive crucial content, such as emergency information, without a subscription.

Unlike existing layer-2 beacon message forwarding schemes and pub-sub-oriented mechanisms, we consider the interest packet naming structure as a candidate for disseminating crucial information among neighboring vehicles. Instead of including a payload for disseminating crucial content or relying on consumer nodes to subscribe, we leverage an interest packet naming structure. Our proposed naming structure contains crucial information within the name of the interest packet. To disseminate the crucial content, we consider the opaque naming structure in NDN, which utilizes interest packets according to the network application requirements. Thus, our proposed push-based content dissemination outperforms existing approaches because this scheme neither contains a payload nor waits for consumers to request specific content.

Secondly, NDN is highly vulnerable to various attacks [12] with severe implications, especially in the context of VNDN, where human lives are directly involved. In VNDN, the illusion of attacker vehicles can disseminate invalid information among their neighboring vehicles. For instance, an illusion attacker vehicle informs its neighboring vehicles about the traffic jam in a particular location. In contrast, there is no traffic jam at the corresponding location. Due to such bogus information dissemination, trust between vehicles can be compromised. Therefore, it is essential to consider the existence of attacker vehicles during crucial content dissemination. They can jeopardize road safety with several attacks, including content pollution attacks (CPA), Distributed Denial of Services (DDoS) attacks [13], and illusion attacks [14].

To identify and prevent these attacks, Machine Learning (ML) plays a significant role in learning from previous behavior and predicting their behavior with high accuracy. Conversely, traditional rule-based systems and other conventional methods are limited by the explicit knowledge programmed into them. In contrast, ML models can uncover and leverage complex patterns and relationships within data that are difficult or impossible to capture with traditional methods. Additionally, ML can handle

large amounts of data, making it well-suited for complex and dynamic problem domains. It can be adapted to new data and situations, making it a flexible and robust approach.

Prior research has made noteworthy contributions to mitigating several attacks using ML techniques. In [15], the Deep Learning (DL) techniques were exploited to detect the misbehavior of vehicles using the Vehicular Reference Misbehavior (VeReMi) dataset [16]. The authors in [17–21] proposed a misbehavior detection system based on ML in VANET. Another work in [22] utilized ML and DL for misbehavior detection in VANET. However, these studies provided learning-based solutions with fewer classifiers, and none of them mentioned above exploited the misbehavior detection system in VNDN.

Unlike the above-mentioned research, we initially address crucial content dissemination using an interest packet naming structure. We utilize the interest packet naming structure to disseminate the crucial content among neighboring vehicles. Secondly, we enable vehicles to identify and prevent illusion attackers in VNDN. To achieve our goal, we evaluate the effectiveness of five ML classifiers and exhibit the best algorithm for detecting illusion attacker vehicles. To the best of our knowledge, this research article presents a novel contribution that exploits interest packet naming structure for disseminating crucial messages among neighboring vehicles in VNDN. Additionally, our ML-based misbehavior detection system identifies and prevents attacker vehicles. The key contributions of this research are as follows:

- We propose push-based crucial content dissemination using an interest packet naming structure without violating the native interest-data architecture of NDN.
- We enable the vehicles to classify the illusion attacker and the legitimate vehicles using ML techniques. By leveraging ML evaluation, vehicles can act or reject crucial information.
- We evaluate the performance of five ML classifiers and identify the most accurate classifier to detect illusion attacker vehicles.

The rest of the article is structured as follows: Section 2 outlines related work in push-based content dissemination and ML-based misbehavior detection in VANET. Section 3 presents a push-based information dissemination architecture. We evaluate the ML classifiers' performance in Section 4. Section 5 exhibits the experimental results. Section 6 concludes the paper. We present the future work in Section 7.

2 Related Work

VNDN has emerged as a promising communication paradigm for Intelligent Transportation Systems (ITS), enabling efficient data dissemination among vehicles and infrastructure. Push-based content dissemination has been proposed to address pull-based content dissemination challenges, and ML techniques have been introduced to detect attacker vehicles in VNDN. This section reviews the related work on push-based crucial content dissemination and ML-oriented misbehavior detection systems in VNDN. We divide the related work into two categories, which are described below.

2.1 Push-Based Content Dissemination

The push-based content dissemination is an excellent addition to NDN, particularly in VANET, where emergency messages cannot wait for consumers to request the content. Although VNDN has been studied over the last decade, push-based techniques are still in the infancy stages of study. Yaqub et al. [23] modified vanilla NDN by introducing a single interest with multiple data packets. They modified the interest packet format with an additional field Crucial Data (cData), which indicates that the consumer requests crucial data. Upon identifying the crucial data request, the

nodes with a 1-hop must store crucial information in PIT until the corresponding PIT entry expires. Another push-based routing mechanism in [11] proposed a hyperbolic forwarding scheme that exploits the betweenness and popularity of nodes. Similar to pub-sub, a hyperbolic push-based mechanism introduces a new table named Pending Data Table (PDT), which stores every pushed interest packet. PDT never expires even after content has been served. This architecture enables the consumer nodes to push interest packets with their coordinates. Based on the PDT, the intermediate node forwards the interest packet to the nearest producers.

Moreover, authors in [24] proposed a push-based critical content dissemination approach using beacon messages to advertise crucial messages among nearby RSUs. Subsequently, neighboring nodes send interest packets to receive the critical content. Similarly, reference [10] presented a push-based beacon message dissemination scheme in VNDN. Another work in [25] proposed a beacon message-oriented emergency message dissemination in VANET. In addition, the pub-sub paradigm has gained significant attention in the literature. Reference [26] introduced a group-based pub-sub content dissemination scheme that enables the nodes to subscribe to the specific content of their interest within a particular group. They proposed a subscription controller that holds various subscription topics. The pub-sub architecture in [27–29] has significantly contributed to pushing infotainment, news, and advertisement services among subscribers. However, this scheme is infeasible for dynamic critical content dissemination, where neighboring nodes must receive crucial content without a subscription. Table 1 summarizes and presents the limitations of existing push-based content dissemination.

Table 1: Summarized push-based related works and their limitations

Reference	Architecture	Scheme	Limitations
Yaqub et al. [23]	VNDN	Request-response	The consumer node must broadcast an interest packet to discover crucial data. This scheme requires an interest packet similar to a pull-based scheme.
Yang et al. [11]	NDN	Betweenness-based push mechanism	The consumer node broadcasts an interest packet containing its coordinates to receive data packets from the nearest node. However, this scheme cannot push a critical message without considering the interest packet.
Yaqub et al. [24]	VNDN	Beacon message-based critical content dissemination	The beacon message contains a payload forwarded from one vehicle to another. However, a flood of beacon messages results in high latency.
Majeed et al. [10]	VNDN	Beacon message-based critical content dissemination	Beacon messages are continuously forwarded until they reach an RSU. This scheme introduces an additional beacon message.

(Continued)

Table 1 (continued)

Reference	Architecture	Scheme	Limitations
Nour et al. [26]	NDN	pub-sub	Producers disseminate specific content among a group of subscribers. However, they cannot dynamically send emergency information to their unsubscribed neighboring vehicles.
Chen et al. [27]	NDN	pub-sub	Due to a specific topic-based subscription scheme, this approach cannot support disseminating dynamically generated crucial messages.
Zhang et al. [28]	NDN	Partial data synchronization (PSync) pub-sub	Sync enables subscribers to receive both partial and complete data. However, it cannot be suitable in dynamic crucial content dissemination.
Patil et al. [29]	NDN	State Vector Sync (SVS) pub-sub	The SVS assigns a sequence number to each data packet and transmits it to the subscriber nodes. However, this scheme is infeasible for crucial content dissemination.
Bilal et al. [25]	VANET	Beacon-oriented message dissemination	This mechanism uses a beacon message to propagate critical information. However, the network architecture in the proposed work differs from NDN.

Unlike the push-based content dissemination mentioned in VNDN, we utilize an interest packet naming structure to deliver crucial information among neighboring vehicles. Our push-based approach can disseminate crucial information without any subscription or payload.

2.2 Misbehavior Detection

Disseminating false information among vehicles in VANETs can result in severe accidents and other safety hazards [30]. It has been a significant concern for academia and industry for many years. The impact of attacks in VANET can be devastating, with disruption to public safety and business activities leading to significant loss of life and property. Existing literature has actively exploited several strategies to cope with attacks, including static threshold-based content identification, prevention, and cryptographic solutions. Authors in [31] exploited ML and DL-based misbehavior detection systems to identify and predict the behavior of vehicles. We present each misbehavior detection technique separately as follows:

2.2.1 Non-Learning-Based Misbehavior Detection

Authors in [32] proposed a reputation-based content identification system in which vehicles evaluate the credibility of content senders before relying on their content. This approach considers

content trustworthy if the sender's reputation meets a specified threshold value. The vehicles then share the reputation with a trusted authority. The authors in reference [33] examined the level of cooperation and selfishness among relay vehicles based on their ability to forward packets to the intended receiver. This architecture considers vehicles cooperative if they meet a specified threshold value, while those without a threshold value are considered selfish. Khelifi et al. [34] proposed a static threshold-based reputation evaluation system for content producer vehicles. Each vehicle has an initial reputation value in this system, updated according to its behavior.

Similarly, Yang et al. [35] designed a blockchain-based reputation system for assessing content credibility in VANET. This approach assigns ratings to the content-providing vehicles and forwards them to a temporarily selected node as blocks. The temporarily selected node is responsible for propagating the reputation among other nodes. In addition, Sun et al. [30] proposed a reputation-based prediction scheme that evaluates the vehicle's reputation using Kalman filtering and chi-squared tests to determine the most reputable vehicle in the network. Other vehicles decide whether to follow a particular vehicle based on its trustworthiness. Table 2 depicts the related works and their limitations in a non-learning-based misbehavior detection system.

Table 2: Summarized non-learning-based related works and their limitations

Reference	Environment	Limitations
Li et al. [36]	VANET	In a static predefined threshold-based reputation system, a specific threshold value is set beforehand, and content senders are deemed trustworthy if their reputation score exceeds this threshold. This approach does not consider the dynamic nature of the network or the fact that reputations may change over time. The lack of granularity may lead to unfairness and discourage users from participating.
Kang et al. [37]	VNDN	
Khelifi et al. [38]	VANET	
Sun et al. [30]	VANET	
Yang et al. [35]	VANET	

2.2.2 ML-Based Misbehavior Detection

ML algorithms have significantly contributed to various fields, including healthcare [39] and anomaly detection [40]. Specifically, they play a critical role in identifying attacks through the binary classification method on datasets, making them an essential tool for attack detection. ML can learn from data automatically and improve its performance over time. The literature has proposed several approaches that leverage the precision of various ML classifiers to detect and mitigate attacks. In [41], the authors presented a Support Vector Machine (SVM) and Logistic Regression (LR) algorithm to detect position falsification attacks in VANET. They exploited the sender's position and speed features. The results indicated that SVM outperformed LR. The authors in [42] proposed a beacon message dissemination system where attackers and legitimate vehicles propagate Basic Safety Messages (BSM) among neighboring vehicles and RSUs. The RSUs accumulate all the BSMs and perform ML evaluation to validate the content based on the previous credibility shared by vehicles. The RSU propagates such information among all the neighboring RSUs and vehicles when it classifies a vehicle as an attacker. Moreover, reference [18] proposed an ML-based misbehavior detection scheme and compared their results with previous works, which demonstrates better performance than others in K-Nearest Neighbor (KNN) and Random Forest (RF). Similarly, the authors in [43] focused on developing ML models to detect misbehaving vehicles. The authors employed binary and multi-classification techniques. Their significant contributions showed promising results for detecting misbehaving vehicles in traditional VANETs [44] designed an Australian Dataset for Misbehaviour Analysis (ADMA) for misbehavior detection in VANET. The authors evaluated their dataset over five

ML classifiers: SVM, KNN, RF, Decision Tree (DT), and Gaussian Naïve Bayes (GNB). According to the results, RF performed an outstanding job compared to other ML classifiers. Table 3 summarizes the related ML-based work and their limitations in the misbehavior detection systems.

Table 3: Summarized ML-based related works and their limitations

Reference	Environment	Limitations
Singh et al. [41]	VANET	The authors assessed the accuracy of only two ML classifiers, SVM and LR.
Sharma et al. [42]	VANET	The proposed approach is limited to BSM and evaluates four ML classifiers.
Ercan et al. [18]	VANET	This work evaluated the accuracy of KNN and RF classifiers. The accuracy score of the proposed classifiers is satisfactory but not highly accurate.
Sonker et al. [43]	VANET	Although the authors achieved high accuracy, they did not exploit ML in VNDN, which differs from traditional VANET.
Amanullah et al. [44]	VANET	The authors evaluated ML classifiers using a BurST-ADMA dataset. This dataset contains a smaller number of messages as compared to the VeReMi dataset.

The above-mentioned related works either proposed a static threshold-based misbehavior detection method or utilized a limited number of ML classifiers to detect the misbehavior of vehicles. Additionally, none of those works focused on misbehavior detection in VNDN. Conversely, we propose five ML algorithms and compare their accuracy in detecting illusion attacks in VNDN.

3 System Model

This section presents our proposed push-based crucial content and regular interest packet dissemination system. Our system model includes the key components and their roles. This section also presents the naming structure for crucial and regular interest packets in this section. Additionally, we propose a reputation dissemination system among RSUs for performing ML classifications. Finally, this section exhibits a detailed algorithm for crucial content dissemination, reputation propagation, and ML classification.

3.1 System Components

3.1.1 Trusted Authority (TA)

We consider the traffic department a trusted authority solely responsible for vehicle registration and the issuance of vehicle identification numbers. It is worth mentioning that the traffic department has no other role except vehicle registration.

3.1.2 Vehicles

Our system consists of OBU-equipped vehicles that are capable of delivering emergency messages. The vehicles can act as consumers, producers, or intermediate (relay) nodes, depending on the

situation. For example, a vehicle that is a consumer in one scenario may act as a producer in another scenario.

3.1.3 RSUs

The RSUs in our system collect the reputation values of the host vehicle, evaluate their authenticity based on their reputation score, and employ ML techniques to identify attackers and benign vehicles. RSUs maintain a database accessible to any vehicle. We assume that all infrastructure-based RSU nodes are connected to a wired network and use a blockchain network to share the reputation among all RSUs.

3.2 Interest Packet Propagation

In NDN, interest packets are structured hierarchically and contain content requests. These packets include the content name and additional information, such as nonce, hop limit, and other optional fields. Unlike traditional NDN, our system comprises two types of interest packets: *crucial interest packets* and *regular interest packets*. The crucial interest packets contain information for other vehicles, such as incident information, and the regular interest packet requests content similar to the NDN interest packet.

3.3 Crucial Information Dissemination

The communication mechanism of our proposed system initiates with transmitting an interest packet among 1-hop neighboring vehicles. The neighboring vehicles first identify the type of interest packet. If it is a crucial message, they check the reputation table to verify the credibility of the host vehicle. The crucial packet is discarded without further action if the host vehicle is identified as an attacker. Conversely, if the host vehicle is deemed legitimate, the neighboring vehicles follow the crucial message. It is important to note that each vehicle can receive an aggregated reputation score from an RSU using a regular interest packet with the optional field "MustBeFresh (MBF)". We distinguish between critical and regular interest packets based on their naming structure, as mentioned below:

3.3.1 Regular Interest Packet Naming Structure

The non-crucial interest packets are pull-based content retrieval packets in VNDN. The consumer requires content, and the producer nodes provide the content as a data packet. The naming structure of a regular interest packet is written as follows:

VNDN/Infotainment/Music/xyz.mp4

Where a vehicle requests music as an infotainment service, the neighboring vehicle checks its CS for a matching interest packet and provides the content to the consumer. If the requested interest packet is not in CS, it is recorded in the PIT table. The PIT records unsatisfied interests and their interface. If the same interest exists in the table, the PIT updates the interface information and drops the duplicate packet. All unsatisfied interest packets are forwarded to the FIB for prefix matching.

3.3.2 Crucial Packet Naming Structure

We place the word *crucial* at the beginning of the naming structure to differentiate crucial packets from regular interest packets. The nearby vehicles instantly recognize it as a crucial packet and act accordingly based on the host vehicles' reputation. According to Algorithm 1, neighboring vehicles will either follow or ignore the crucial packet based on the reputation of the host vehicle. If the host

vehicle is identified as an attacker, the packet will be dropped, and no action will be taken. However, if the host vehicle is recognized as legitimate, neighboring vehicles will act upon the message. The crucial messages are no longer required to be stored in the PIT of neighboring vehicles. The naming structure of the crucial message is:

Vehicle_No/Crucial/Incident/location/traffic_congestion/avoid

Where a vehicle notifies other vehicles to avoid a specific location due to heavy traffic, in this situation, the attacker vehicles can manipulate their exact location with a fake position.

3.4 Reputation Propagation

After verifying the legitimacy of the content provided by the host vehicle, neighboring vehicles generate a reputation (either positive (0) or negative (1)) and forward it to the RSUs using a reputation dissemination scheme similar to the crucial packet dissemination mechanism mentioned above. The naming structure for the reputation dissemination scheme is as follows:

Reputation/VNDN/Vehicle/No/1 or 0

We assign (1) to illusion attackers and (0) to legitimate vehicles. The term "reputation" refers to the reputation of the host vehicle, in which RSUs accept the packets and vehicles ignore them. Afterward, the RSU incorporates all the reputations of a host vehicle and performs ML algorithms to classify the vehicle. Table 4 depicts the notation used in Algorithm 1, and Fig. 2 shows an overall transmission and ML performance evaluation architecture.

Table 4: Summary of notations used in Algorithm 1

Notation	Description
I_{PKT}	Interest packet
D_{PKT}	Data packet
C_{PKT}	Crucial packet
CP_R	Content producer reputation
A_T	Attacker
Legit	Legitimate
Inv	Invalid
Norm	Normal
Cnt	Content

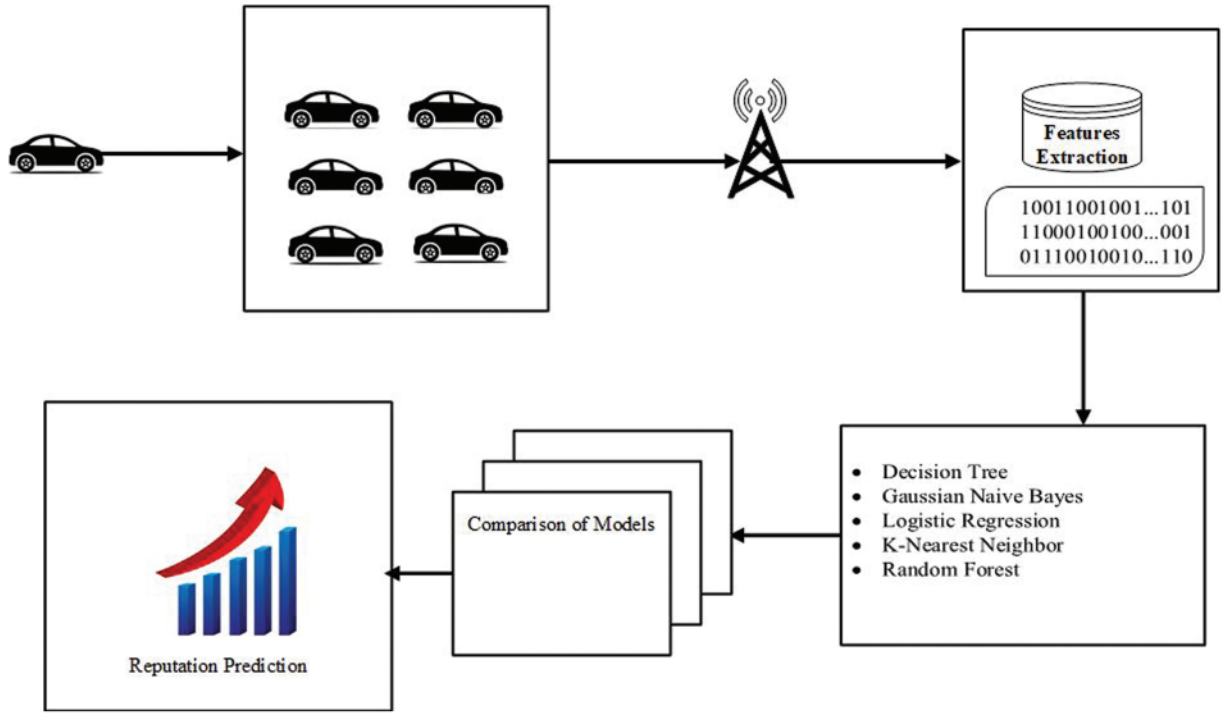

Figure 2: Crucial message dissemination and ML-based reputation verification

Algorithm 1: Interest packet identification and forwarding mechanism

Require: I_{PKT}
1: **if** $I_{PKT} = C_{PKT}$ **then**
2: verify the reputation of the host
3: **if** $CP_R = A_T$ **then**
4: Drop packet
5: **else if** $CP_R =$ Legit **then**
6: Validate content
7: **end if**
8: **if** $CP_R =$ legit **then**
9: Assign 0 to the host
10: push reputation to RSUs
11: **else if** $CP_R =$ Inv **then**
12: Assign 1 to the host
13: push reputation to RSUs
14: **end if**
15: **else if** $I_{PKT} =$ Norm **then**
16: **if** Cnt $\in$ CS **then**
17: Create a Data Packet
18: Send data packet to the incoming interface
19: **end if**
20: **if** Cnt $\in$ PIT **then**
21: Add interface

(Continued)

Algorithm 1 (Continued)

```
            remove the interest packet
        else
            Add entry in the PIT
            Forwarded Interest to FIB
        end if
end if
```

4 ML-Based Evaluation

This section evaluates the behavior of vehicles and classifies them as benign or attacker vehicles using the ML-based binary classification technique. The ML evaluation comprises three stages: dataset collection and extraction, data preparation, and binary classification.

4.1 Dataset Collection and Extraction

Our ML-based misbehavior identification process uses an open dataset called VeReMi [16], designed for detecting vehicles' misbehavior. The VeReMi simulation is performed on OMNET++ and Simulation of Urban Mobility (SUMO) [45] on the Luxembourg traffic scenario (LuST) [46]. It consists of 255 simulations with heterogeneous traffic scenarios, including five different types of attacks: constant (type 1), constant offset (type 2), random (type 4), random offset (type 8), and eventual stop (type 16). The description of these attacks is depicted in Table 5. The simulation contains three attacker densities (10%, 20%, and 30%) with multiple log files and a single truth file in each simulation. The log files contain messages received from multiple vehicles, including legitimate messages (0) and attack messages (1).

Table 5: VeReMi attack types

Attacker type	Description
Type 1: Constant	The attacker vehicle transmits a fixed location to neighboring vehicles.
Type 2: Constant offset	The attacker injects a fixed offset to the original position.
Type 4: Random	The attacker injects random coordinates into its propagated messages.
Type 8: Random offset	The attacker vehicle sends random positions from a pre-configured position around the vehicle.
Type 16: Eventual stop	The attacker initially pretends to be a benign vehicle and transmits the same position repeatedly.

4.2 Data Preparation

After determining the dataset, we converted all the log files from JSON to CSV format and merged them with ground truth files to generate a labeled dataset for each attack type. We removed duplicate messages and non-contributing features such as position noise, speed noise, and RSSI. To train and test our model efficiently and to prevent over-fitting, we performed k-fold cross-validation. We split the entire dataset into k-fold and applied 10-fold cross-validation.

4.3 Classification

We utilized binary classification to distinguish between attacker and legitimate vehicles across five attack environments of VeReMi. We experimented with various ML classifiers, including DT, KNN, RF, GNB, and LR.

4.4 Performance Evaluation

We trained and tested various ML classifiers on the VeReMi dataset to see how well they work. We analyzed the performance in terms of accuracy, precision, recall, and the F-1 score:

4.4.1 Accuracy

It is a ratio of correct detections, True Positive (TP) and True Negative (TN), out of the total number of detections: TP, TN, False Positive (FP), and False Negative (FN). The accuracy is measured as shown in Eq. (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

4.4.2 Precision

The proportion of misbehavior was accurately identified out of all values. A lower precision value shows that the model has many FP. Eq. (2) shows the precision model.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

4.4.3 Recall

It is described as the proportion of misbehavior accurately identified as actual misbehavior. The lower recall score demonstrates a model's inability to identify misbehavior. Eq. (3) refers recall model.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

4.4.4 F-1 Score

As the harmonic mean of recall and precision, the F-1 score evaluates the overall detection quality of a model. Eq. (4) shows the F-1 model.

$$F{-}1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

5 Experimental Results and Discussion

Our research aims to identify and mitigate illusion attacks in VNDN. To achieve this, we evaluate the performance of various ML algorithms based on the behavior of vehicles. In order to perform ML classification, we implemented our ML classification in Python using an Intel desktop core i7 CPU with 2.8 GHz and 16 GB of RAM. We conducted our experiments on the Windows 11 operating system. To evaluate the effectiveness of each ML classifier in detecting vehicle misbehavior, we utilized a precision-recall curve [47] in conjunction with a Receiver Operating Characteristic (ROC) curve to visualize the outcomes. Precision and recall curves are commonly employed for assessing binary classification performance. The curve indicates the trade-off between precision and recall values, where a larger area under the curve indicates higher values for both metrics. A high precision value

corresponds to a low false positive rate, whereas a high recall value corresponds to a low false negative rate. As shown in Table 6, our evaluation involves assessing the performance of various ML classifiers on different attacker scenarios using the VeReMi dataset. Based on the results, RF and DT have shown excellent performance in detecting all attacks. Similarly, KNN has also achieved better results. On the other hand, GNB and LR showed poor results. The performance of each ML classifier for each attack is mentioned below.

Table 6: ML classifiers' performance evaluation

ML classifiers	Performance measures	Type 1	Type 2	Type 4	Type 8	Type 16
DT	Precision	1.00	1.00	0.97	0.97	0.98
	Recall	0.99	0.99	0.86	0.85	0.87
	Accuracy	0.99	0.99	0.91	0.91	0.93
	F-1 score	0.99	0.99	0.91	0.91	0.93
KNN	Precision	0.99	0.98	0.86	0.86	0.90
	Recall	0.85	0.83	0.73	0.73	0.75
	Accuracy	0.92	0.91	0.81	0.80	0.83
	F-1 score	0.91	0.90	0.79	0.82	0.84
GNB	Precision	0.54	0.54	0.51	0.51	0.52
	Recall	0.41	0.35	0.51	0.62	0.52
	Accuracy	0.53	0.53	0.51	0.52	0.52
	F-1 score	0.46	0.42	0.51	0.52	0.52
RF	**Precision**	**1.00**	**1.00**	**0.97**	**0.97**	**0.98**
	Recall	0.99	0.99	0.85	0.85	0.86
	Accuracy	1.00	1.00	0.91	0.91	0.92
	F-1 score	1.00	1.00	0.90	0.90	0.92
LR	Precision	0.53	0.53	0.51	0.51	0.51
	Recall	0.53	0.50	0.55	0.54	0.55
	Accuracy	0.53	0.53	0.51	0.51	0.51
	F-1 score	0.53	0.52	0.51	0.53	0.53

Attack Type 1: It is relatively simple to identify attackers due to constant position transmission with unfixed velocity. As shown in Fig. 3, RF, DT, and KNN reflect outstanding performance in terms of precision, recall, and F-1, whereas GNB and LR demonstrate poor performance.

Attack Type 2: The attackers continuously broadcast a constant offset position using their actual position in this attack. Like attack type 1, the RF, DT, and KNN showed outstanding accuracy in attack type 2. However, GNB and LR indicated a lower performance. Fig. 4 illustrates the performance of all ML classifiers in detecting attack type 2.

Attack Type 4: The type 4 attackers constantly broadcast a new random position for each message. The ML classifiers such as RF, DT, and KNN exhibit high accuracy in identifying the attacks, as shown in Fig. 5. Moreover, LR and GNB reflected unsatisfactory performance.

Attack Type 8: In attack type 8, the attackers broadcast a pre-configured random position for each content. The results depicted in Fig. 6 show that RF achieves 95% accuracy, while DT and KNN achieved satisfactory results. On the other hand, LR and GNB have a low accuracy performance.

Attack Type 16: The attackers initially transmit their actual position, similar to a benign vehicle, but later turn to a stable position. This type of attack is challenging to detect due to its floating behavior, making it difficult to identify the misbehavior of attackers. Fig. 7 shows the performance of various ML classifiers in detecting attack type 16. The results indicate that the RF classifier achieved higher accuracy than the other ML algorithms.

We evaluated the performance of various ML classifiers on the VeReMi dataset. Our observations revealed that RF and DT classifiers achieved high accuracy rates in terms of precision, recall, and F-1 score across all the attack types. Specifically, the RF classifier delivered an outstanding performance, achieving 100% accuracy in attack types 1 and 2 and over 95% accuracy in the other attack types. Table 6 presents the performance evaluation results of the ML classifiers on the VeReMi dataset. We observed that DT provided highly accurate results, achieving 99% accuracy in attack types 1 and 2 and more than 91% accuracy in other attack types. Similarly, KNN also showed promising results, with 96% accuracy in attack types 1 and 2, 89% accuracy in attack types 4 and 8 and 90% accuracy in attack type 16. In contrast, LR and GNB demonstrated poor performance. Our findings suggest that RF and DT can effectively detect illusion attacker vehicles accurately.

Visualized Results

We employed various ML classifiers to the VeReMi dataset and compared their efficiency based on obtained results. To assess the performance of our ML models, we visualized our results using accuracy calculation techniques as mentioned in Eqs. (1)–(4). First, we utilized precision-recall and ROC curves to evaluate the trade-off between precision and recall for binary classification. The Area Under Curve (AUC) curves indicated strong performance in three models, with the highest AUC achieved by the RF, DT, and KNN models. In contrast, GNB and LR depicted poor performance. Based on the experimental findings, RF and DT accurately classify illusion attackers and legitimate vehicles. These classifiers effectively detected the misbehavior of vehicles in all attack types. Overall, our visualized results provide a comprehensive view of the performance of our ML models. From Figs. 3–7, we can observe that RF and DT perform well in five different attacker scenarios.

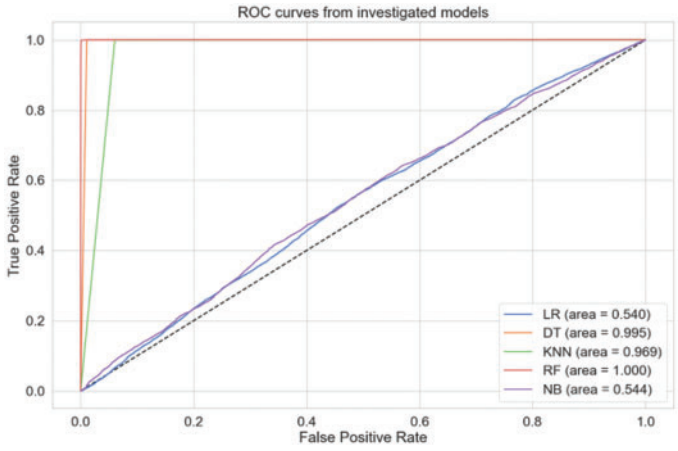

Figure 3: Attack type 1 results

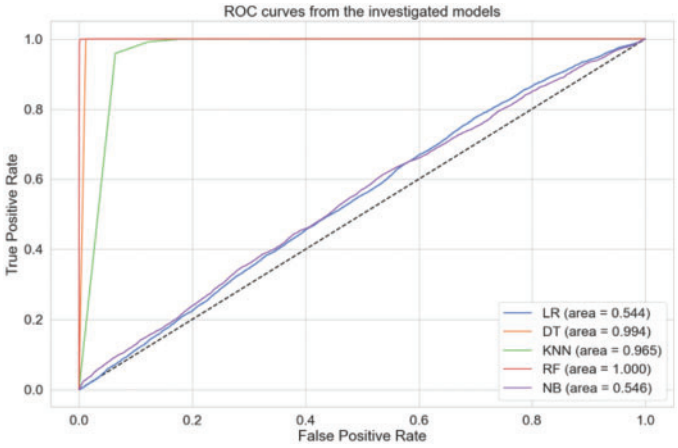

Figure 4: Attack type 2 results

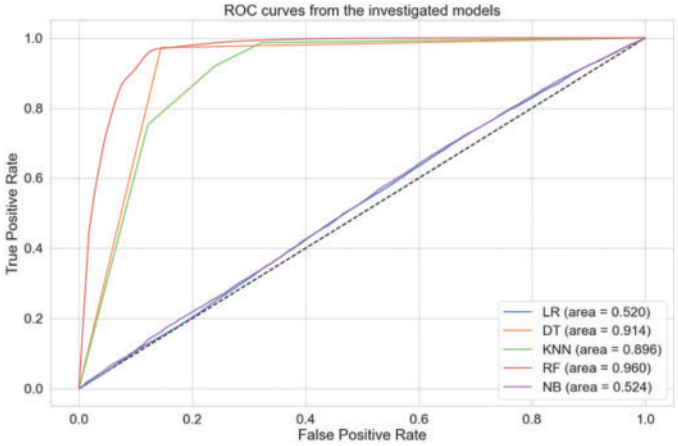

Figure 5: Attack type 4 results

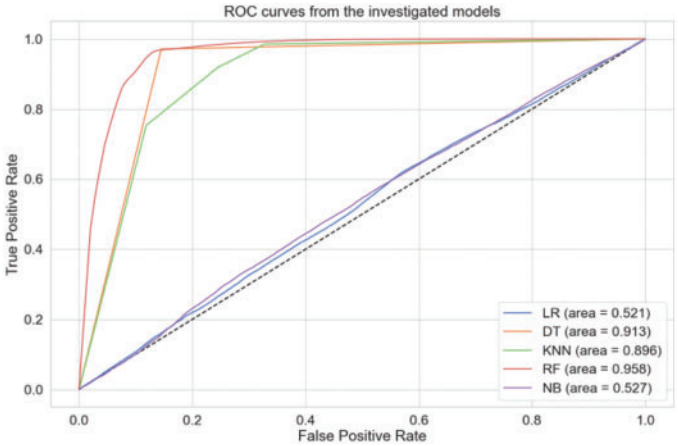

Figure 6: Attack type 8 results

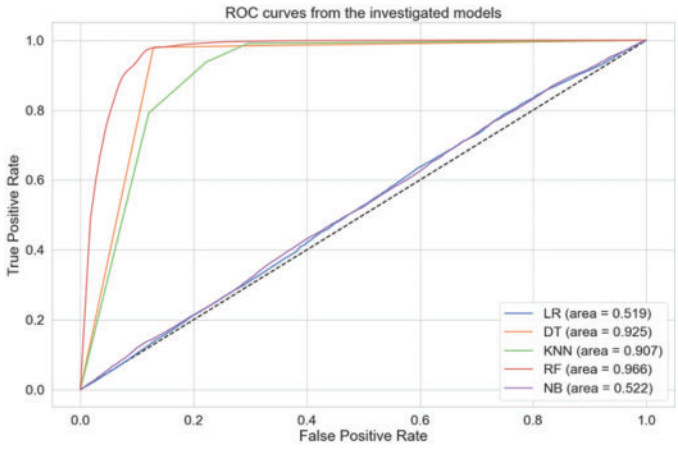

Figure 7: Attack type 16 results

6 Conclusion

The crucial information dissemination and its trustworthiness are the essential facets of VNDN due to the involvement of human lives in it. Fake information dissemination can lead to road accidents, traffic congestion, and increased energy consumption. In addition, the traditional NDN in VANET lacks push-based content dissemination and fake information detection techniques. In order to address these challenges, this research aims to improve NDN's capability to distribute critical messages in the VNDN environment without relying on an interest packet. In this connection, we used a novel push-based content dissemination technique in VNDN. We leveraged the NDN's interest packet naming structure to disseminate crucial content without waiting for a content consumer to broadcast an interest packet. However, push-based content dissemination can be vulnerable to illusion attacks where adversaries can inject fake information and broadcast it among neighboring vehicles. To cope with this risk, we enabled neighboring vehicles to query the aggregate reputation of host vehicles from RSUs. The RSUs classify host vehicles as attackers or legitimate based on ML classification. This study used five ML classifiers, including DT, RF, KNN, GNB, and LR, on a publicly available dataset VeReMi. Our proposed ML classification results demonstrate that RF and DT accurately detect all attack types within the VeReMi dataset. RF achieved an accuracy of 100% in attack types 1 and 2. It achieved 95% accuracy in attack types 4 and 8 and 96% in attack type 16. On the other hand, DT achieved an accuracy of 99% in attack types 1 and 2, whereas it attained 91% accuracy in attack types 4 and 8 and 93% in attack type 16. Our research is the first to modify interest packet naming structures and utilize them for crucial content dissemination. Moreover, evaluating ML classifiers for detecting illusion attacks in VNDN advocates the novelty of our proposed study. Thus, our proposed network architecture ensures crucial content dissemination and enhances trust between vehicles.

7 Future Work

Our proposed research classifies attacker and legitimate vehicles using ML classifiers. However, further investigations can focus on exploiting DL classifiers to improve the accuracy and efficiency of the misbehavior detection system in VNDN.

Acknowledgement: The authors acknowledge the Researchers Supporting Project Number (RSP2023R34), King Saud University, Riyadh, Saudi Arabia.

Funding Statement: This work was supported by the Researchers Supporting Project Number (RSP2023R34), King Saud University, Riyadh, Saudi Arabia.

Author Contributions: Study conception and design: A. H. Magsi, G. Muhammad, Z. Ali; data collection: A. H. Magsi, S. Kareem; analysis and interpretation of results: G. Muhammad, S. Memon; draft manuscript preparation: A. H. Magsi, G. Muhammad, Z. Ali, S. Kareem; data curation: S. Memon, S. Kareem, Z. Ali; visualization: A. H. Magsi, G. Muhammad, S. Memon, S. Kareem; resources: S. Kareem, S. Memon; Z. Ali. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used to support the findings of this study are available from authors upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. Jia, K. Lu, J. Wang, X. Zhang and X. Shen, "A survey on platoon-based vehicular cyber-physical systems," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 263–284, 2015.

[2] L. Sun, Z. Lin, W. Li and Y. Xiang, "Freeway incident detection based on set theory and short-range communication," *The International Journal of Transportation Research*, vol. 11, no. 10, pp. 558–569, 2019.

[3] S. Al-Sultan, M. M. Al-Doori, A. H. Al-Bayatti and H. Zedan, "A comprehensive survey on vehicular ad hoc network," *Journal of Network and Computer Applications*, vol. 37, no. 10, pp. 380–392, 2014.

[4] K. C. Dey, A. Rayamajhi, M. Chowdhury, P. Bhavsar and J. Martin, "Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network—performance evaluation," *Transportation Research Part C: Emerging Technologies*, vol. 68, no. 6, pp. 168–184, 2016.

[5] H. Wang, S. Adhatarao, M. Arumaithurai and X. Fu, "COPSS-lite: Lightweight icn based pub/sub for iot environments," arXiv preprint arXiv: 1706.03695, 2017.

[6] C. Chen, C. Wang, T. Qiu, M. Atiquzzaman and D. O. Wu, "Caching in vehicular named data networking: Architecture, schemes and future directions," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 4, pp. 2378–2407, 2020.

[7] H. Khelifi, S. Luo, B. Nour, H. Moungla, Y. Faheem *et al.,* "Named data networking in vehicular ad hoc networks: State-of-the-art and challenges," *IEEE Communications Surveys and Tutorials*, vol. 22, no. 1, pp. 320–351, 2020.

[8] A. Afanasyev, P. Mahadevan, I. Moiseenko, E. Uzun and L. Zhang, "Interest flooding attack and countermeasures in named data networking," in *Proc. of 2013 IFIP Networking Conf.*, Brooklyn, NY, USA, pp. 1–9, 2013.

[9] C. Fang, H. Yao, Z. Wang, W. Wu, X. Jin *et al.,* "A survey of mobile information-centric networking: Research issues and challenges," *IEEE Communications. Surveys & Tutorials*, vol. 20, no. 3, pp. 2353–2371, 2018.

[10] M. F. Majeed, S. H. Ahmed and M. N. Dailey, "Enabling push-based critical data forwarding in vehicular named data networks," *IEEE Communication Letters*, vol. 21, no. 4, pp. 873–876, 2017.

[11] W. Yang, Y. Qin, Z. Yi and Y. Yang, "Content-based hyperbolic routing and push mechanism in named data networking," in *Proc. of IEEE International Conf. on Communications*., Shanghai, China, pp. 1–6, 2019.

[12] A. Hidouri, M. Hadded, H. Touati, N. Hajlaoui and P. Muhlethaler, "Attacks, detection mechanisms and their limits in named data networking (ndn)," in *Proc. of ICCSA*, Malaga, Spain, pp. 310–323, 2022.

[13] S. Signorello, M. R. Palattella and L. A. Grieco, "Security challenges in future NDN-Enabled VANETs," in *Proc. of IEEE Trustcom/BigDataSE/ISPA*, Tianjin, China, pp. 1771–1775, 2016.
[14] H. Khelifi, S. Luo, B. Nour and S. C. Shah, "Security and privacy issues in vehicular named data networks: An overview," *Mobile Information System*, vol. 2018, no. 5, pp. 11, 2018.
[15] X. Liu, "Misbehavior detection based on deep learning for vanets," in *Proc. of CNCIT*, Beijing, China, pp. 122–128, 2022.
[16] R. W. van der Heijden, T. Lukaseder and F. Kargl, "VeReMi: A dataset for comparable evaluation of misbehavior detection in vanets," in *Proc. of SecureComm*, Singapore, vol. 254, pp. 318–337, 2018.
[17] S. Gyawali, Y. Qian and R. Q. Hu, "Machine learning and reputation based misbehavior detection in vehicular communication networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8871–8885, 2020.
[18] S. Ercan, M. Ayaida and N. Messai, "Misbehavior detection for position falsification attacks in vanets using machine learning," *IEEE Access*, vol. 10, pp. 1893–1904, 2022.
[19] S. Ercan, M. Ayaida and N. Messai, "New features for position falsification detection in vanets using machine learning," in *Proc. of IEEE ICC*, Montreal, QC, Canada, pp. 1–6, 2021.
[20] A. Khot and M. Dave, "Position falsification misbehavior detection in VANETs," in *Proc. of MRCN*, Haryana, India, pp. 487–499, 2020.
[21] F. Hawlader, A. Boualouache, S. Faye and T. Engel, "Intelligent misbehavior detection system for detecting false position attacks in vehicular networks," in *Proc. of IEEE International Conf. on Communications Workshops*, Montreal, QC, Canada, pp. 1–6, 2021.
[22] R. Sultana, J. Grover, J. Meghwal and M. Tripathi, "Exploiting machine learning and deep learning models for misbehavior detection in VANET," *International Journal of Computers and Applications*, vol. 44, no. 11, pp. 1024–1038, 2022.
[23] M. A. Yaqub, S. H. Ahmed, S. H. Bouk and D. Kim, "Enabling critical content dissemination in vehicular named data networks," in *Proc. of RACS*, New York, USA, pp. 94–99, 2018.
[24] M. A. Yaqub, S. H. Ahmed and D. Kim, "A detailed simulation study of the push-based protocol for critical data dissemination in vehicular named data networks," in *Proc. of NaNA*, Daegu, South Korea, pp. 191–195, 2019.
[25] M. Bilal, E. U. Munir and A. Ullah, "BEMD: Beacon-oriented emergency message dissemination scheme for highways," *Ad Hoc Networks*, vol. 142, no. 7, pp. 103095, 2023.
[26] B. Nour, K. Sharif, F. Li, S. Yang, H. Moungla *et al.,* "ICN publisher-subscriber models: Challenges and group-based communication," *IEEE Network*, vol. 33, no. 6, pp. 156–163, 2019.
[27] J. Chen, M. Arumaithurai, L. Jiao, X. Fu and K. K. Ramakrishnan, "COPSS: An efficient content oriented publish/subscribe system," in *Proc. of ACM/IEEE Seventh Symp. on Architectures for Networking and Communications Systems*, Brooklyn, NY, USA, pp. 99–110, 2011.
[28] M. Zhang, V. Lehman and L. Wang, "Scalable name-based data synchronization for named data networking," in *Proc. of IEEE INFOCOM*, Atlanta, GA, USA, pp. 1–9, 2017.
[29] V. Patil, P. Moll and L. Zhang, "Supporting pub/sub over NDN sync," in *Proc. of 8th ACM Conf. on ICN*, New York, USA, pp. 133–135, 2019.
[30] M. Sun, M. Li and R. Gerdes, "A data trust framework for vanets enabling false data detection and secure vehicle tracking," in *IEEE of Conf. on CNS*, Las Vegas, NY, USA, pp. 1–9, 2017.
[31] P. Sharma, J. Petit and H. Liu, "Pearson correlation analysis to detect misbehavior in VANET," in *Proc. of IEEE 88th VTC-Fall*, Chicago, IL, USA, pp. 1–5, 2018.
[32] Q. Li, A. Malip, K. M. Martin, S. L. Ng and J. Zhang, "A reputation-based announcement scheme for VANETs," *IEEE Transaction on Vehicular Technology*, vol. 61, no. 9, pp. 4095–4108, 2012.
[33] J. A. F. F. Dias, J. J. P. C. Rodrigues, L. Shu and S. Ullah, "Performance evaluation of a cooperative reputation system for vehicular delay-tolerant networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, pp. 1–13, 2014.
[34] H. Khelifi, S. Luo, B. Nour, H. Moungla, S. H. Ahmed *et al.,* "A blockchain-based architecture for secure vehicular named data networks," *Computers and Electrical Engineering*, vol. 86, no. 4, pp. 106715, 2020.

[35] Z. Yang, K. Zheng, K. Yang and V. C. M. Leung, "A blockchain-based reputation system for data credibility assessment in vehicular networks," in *IEEE 28th Annual Int. Symp. on Personal, Indoor, and Mobile Radio Communications (PIMRC)*., Montreal, QC, Canada, PIMRC, pp. 1–5, 2018.
[36] Q. Li, A. Malip, K. M. Martin, S. L. Ng and J. Zhang, "A reputation-based announcement scheme for VANETs," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 9, pp. 4095–4108, 2012.
[37] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan *et al.,* "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4660–4670, 2019.
[38] H. Khelifi, S. Luo, B. Nour, H. Moungla, S. H. Ahmed *et al.,* "A blockchain-based architecture for secure vehicular named data networks," *Computers & Electrical Engineering*, vol. 86, no. 4, pp. 106715, 2020.
[39] G. Muhammad, F. Alshehri, F. Karray, A. El Saddik, M. Alsulaiman *et al.,* "A comprehensive survey on multimodal medical signals fusion for smart healthcare systems," *Information Fusion*, vol. 76, no. 8, pp. 355–375, 2021.
[40] M. M. Islam, S. Nooruddin, F. Karray and G. Muhammad, "Human activity recognition using tools of convolutional neural networks: A state of the art review, data sets, challenges, and future prospects," *Computers in Biology and Medicine*, vol. 149, pp. 106060, 2022.
[41] P. K. Singh, S. Gupta, R. Vashistha, S. K. Nandi and S. Nandi, "Machine learning based approach to detect position falsification attack in VANETs," in *Proc. of ICEA-ISAP*, Jaipur, India, pp. 166–178, 2019.
[42] A. Sharma and A. Jaekel, "Machine learning based misbehaviour detection in VANET using consecutive BSM approach," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 1–14, 2021.
[43] A. Sonker and R. K. Gupta, "A new procedure for misbehavior detection in vehicular ad-hoc networks using machine learning," *International Journal of Electrical and Computing Engineering*, vol. 11, no. 3, pp. 2535–2547, 2021.
[44] M. A. Amanullah, M. B. Chhetri, S. W. Loke and R. Doss, "BurST-ADMA: Towards an Australian dataset for misbehaviour detection in the internet of vehicles," in *Proc. of PerCom Workshops*, Pisa, Italy, pp. 624–629, 2022.
[45] M. Behrisch, L. Bieker, J. Erdmann and D. Krajzewicz, "SUMO-simulation of urban mobility: An overview," in *Proc. of SIMUL*, Barcelona, Spain, pp. 23–28, 2011.
[46] L. Codeca, R. Frank, S. Faye and T. Engel, "Luxembourg sumo traffic (lust) scenario: Traffic demand evaluation," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 52–63, 2017.
[47] K. Boyd, K. H. Eng and C. D. Page, "Area under the precision-recall curve: Point estimates and confidence intervals," in *ECML PKDD*, Prague, Czech Republic, pp. 451–466, 2013.