Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2023.040290

ARTICLE

A Machine Learning-Based Attack Detection and Prevention System in Vehicular Named Data Networking

Arif Hussain Magsi[1,*], Ali Ghulam[2], Saifullah Memon[1], Khalid Javeed[3], Musaed Alhussein[4] and Imad Rida[5]

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

[2]Information Technology Center, Sindh Agriculture University, Tandojam, 70050, Pakistan

[3]Department of Computer Engineering, College of Computing and Informatics, University of Sharjah, Sharjah, 27272, United Arab Emirate

[4]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, P. O. Box 51178, Riyadh, 11543, Saudi Arabia

[5]Laboratory Biomechanics and Bioengineering, University of Technology of Compiegne, Compiegne, 60200, France

*Corresponding Author: Arif Hussain Magsi. Email: ahmagsi@bupt.edu.cn

Received: 13 March 2023 Accepted: 13 June 2023 Published: 29 November 2023

ABSTRACT

Named Data Networking (NDN) is gaining a significant attention in Vehicular Ad-hoc Networks (VANET) due to its in-network content caching, name-based routing, and mobility-supporting characteristics. Nevertheless, existing NDN faces three significant challenges, including security, privacy, and routing. In particular, security attacks, such as Content Poisoning Attacks (CPA), can jeopardize legitimate vehicles with malicious content. For instance, attacker host vehicles can serve consumers with invalid information, which has dire consequences, including road accidents. In such a situation, trust in the content-providing vehicles brings a new challenge. On the other hand, ensuring privacy and preventing unauthorized access in vehicular (VNDN) is another challenge. Moreover, NDN's pull-based content retrieval mechanism is inefficient for delivering emergency messages in VNDN. In this connection, our contribution is threefold. Unlike existing rule-based reputation evaluation, we propose a Machine Learning (ML)-based reputation evaluation mechanism that identifies CPA attackers and legitimate nodes. Based on ML evaluation results, vehicles accept or discard served content. Secondly, we exploit a decentralized blockchain system to ensure vehicles' privacy by maintaining their information in a secure digital ledger. Finally, we improve the default routing mechanism of VNDN from pull to a push-based content dissemination using Publish-Subscribe (Pub-Sub) approach. We implemented and evaluated our ML-based classification model on a publicly accessible BurST-Asutralian dataset for Misbehavior Detection (BurST-ADMA). We used five (05) hybrid ML classifiers, including Logistic Regression, Decision Tree, K-Nearest Neighbors, Random Forest, and Gaussian Naive Bayes. The qualitative results indicate that Random Forest has achieved the highest average accuracy rate of 100%. Our proposed research offers the most accurate solution to detect CPA in VNDN for safe, secure, and reliable vehicle communication.

KEYWORDS

Named data networking; vehicular networks; reputation; caching; machine-learning

1 Introduction

The persistent increase in traditional vehicles has led to greater convenience for people but has also raised the risk of fatal incidents. As per a forecast made in 2015, the number of automobiles is expected to double in coming 10 to 20 years [1]. In such a situation, VANET [2] is an indispensable and most prominent state-of-the-art solution. It exchanges massive amounts of information to improve transportation efficiency, prevent road accidents, deliver emergency services, provide traffic and weather updates, reduce traffic congestion, and delivers infotainment services. VANET-aided vehicles are equipped with robust Onboard Unit (OBU) resources, including fast computational power, ample storage capacity, and communication capabilities via Dedicated short-range communication (DSRC) and Wireless in Vehicular Environment (WAVE), also denoted as IEEE 802.11p [3]. Within VANET, vehicles can communicate in various ways, including Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication. Despite its numerous capabilities, the existing IP-based communication system in VANET cannot support an extremely complex network caused by the mobility of vehicles, intermittent connectivity, device heterogeneity, and rapid topology changes that cause performance bottlenecks.

To address these unprecedented challenges, Named Data Networking (NDN) [4], which is one of the most effective network paradigms in the realm of Information-Centric Networking (ICN) [5] architecture. NDN is poised to replace the host-centric content transmission with a content-centric network. Van Jacobson is credited for introducing the NDN technology as part of a research project initiated by the Palo Alto Research Center (PARC) [6]. NDN has emerged as a prominent communication architecture in the realm of ICN by leveraging content as the fundamental unit of communication without regard for the content provider or its physical location. NDN comprises three distinct nodes. **Content Consumer:** The content consumer is a resource-intensive node that requests the data. **Content Producer:** It provides the requested data to the content consumer nodes. **Intermediate Node:** The intermediate nodes serve as relay nodes. They forward interest and data packets between consumers and producers. Another vital responsibility of this node involves storing the content within its local cache memory, thereby serving consumers without approaching the producers. The communication in NDN comprises two types of packets: *interest packets* and *data packets.* A content consumer initiates the communication by sending an interest packet that expresses the desired content's name. The intermediate node forwards the interest packet to neighboring nodes if it has no content in its CS. Upon receiving the interest packet, the producer sends a corresponding data packet with the requested content, following the reverse path to the consumer. The data packet contains the content metadata, such as the name or identifier of the content and cryptographic signatures. This interest-data packet exchange model in NDN differs from the traditional client-server model used in IP networks, where a client sends a request to a specific server for content.

In addition, NDN incorporates three essential data structures in each node. **Content Store (CS):** Vehicles can cache content they receive and use it to respond to future requests for the same content. This caching behavior can reduce the network's load and latency by allowing content to be served locally instead of being fetched from the source every time. **Pending Interest Table (PIT):** It is a table in every NDN node that records pending interests and their names received from different interfaces that have yet to be satisfied. **Forward Information Base (FIB):** Maintains information about corresponding outgoing interfaces to route the interest packet.

These data structures allow nodes to efficiently store, retrieve, and route content in the network. The content retrieval mechanism in NDN is based on the pull model, where a consumer node initiates communication by sending an interest packet. Upon receiving the interest packet, the nodes within

one hop of the consumer check their CS for the requested content. If the content is found at any neighboring node, it is returned to the consumer as a data packet. If the requested content is unavailable in CS, the intermediate node turns to the PIT table to check for unmet interests and their interfaces. If a similar interest already exists in the PIT, the interest is discarded, and the interface name is recorded. If no matching interest exists in the PIT, a new entry is created, and the interest is forwarded to the upstream node based on the longest prefix match in FIB. Eventually, the content-producing node provides the required content as a data packet since the provider cannot provide the content without receiving an interest packet. Fig. 1 illustrates the communication architecture of VNDN.

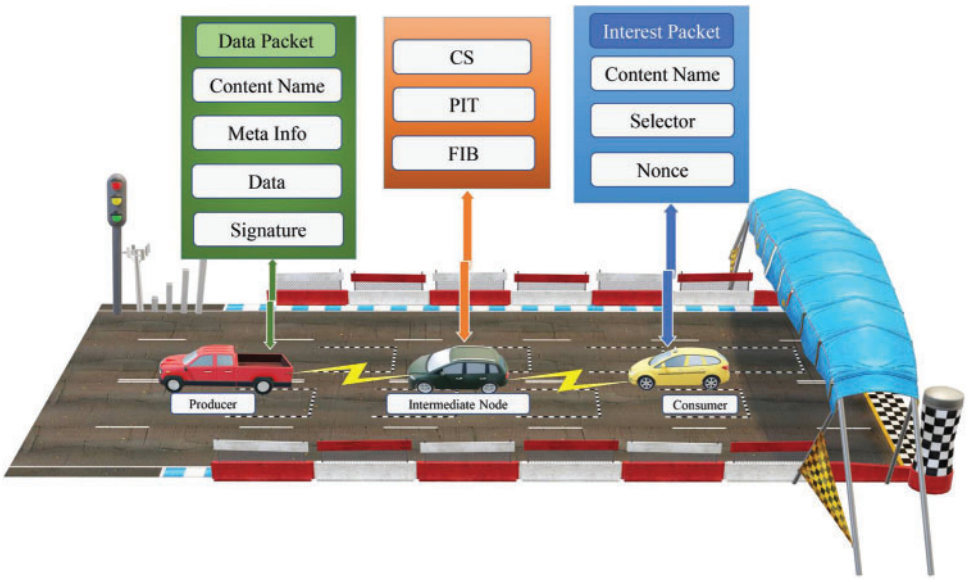

Figure 1: Content transmission in VNDN

NDN has been exploited in various research arenas, namely VANET [7], Internet of Things (IoT) [8], Wireless Sensor Networks (WSN) [9], and others. Unlike IP-based architecture, NDN supports high mobility and a dynamic typology that allows content transmission without establishing a connection between two endpoints. In contrast to IP-based communication, NDN retrieves content using a unique hierarchical name, irrespective of the content provider or its location, which shifted the content retrieval mechanism from "where" to "what" [10]. In particular, NDN has excellent potential to improve network performance by enhancing content access and reducing communication delay.

However, existing VNDN encounters various challenges, including caching, routing, security, and privacy. In particular, security attacks, including Content Poisoning Attacks (CPA) [11], Interest Flooding Attacks (IFA) [12] and Man-in-Middle Attacks [13]. These attacks can pose significant threats to network security and should be mitigated through appropriate measures. Specifically, CPA is exceptionally challenging due to attacker vehicles employing forged packets with correct names. Consequently, intermediate nodes receive, cache and serve the malicious content to consumers. Thus, the whole network fails to receive the correct information. On the other hand, the default content caching strategy in NDN, such as Leave Copy Everywhere (LCE), caches and forwards the content among other nodes without evaluating the legitimacy of content, which leads the intermediate nodes to store a plethora of incorrect content in their local storage.

Another drawback of the existing NDN in VANET is its default receiver driven content dissemination mechanism, where vehicles must initial an interest packet to receive corresponding information. However, this request-response mechanism fails to provide incident reports, traffic information and advertisements in VANET. In VANET, the vehicles cannot wait for content consumers to request specific information. For instance, traffic jam information in a specific location must be shared among other vehicles without expecting an interest packet.

Meanwhile, existing NDN in VANET lacks to ensure the vehicles' privacy due to the unavailability of reliable and secure infrastructure. The centralized cloud architecture [14] is a famous infrastructure to provide several services. For example, companies like Amazon and eBay leverage a trusted third party to maintain their reputations [15]. However, centralized cloud infrastructure is infeasible in VNDN due to the mobility of vehicles.

In order to address the issues mentioned above, several innovative methodologies have been suggested in the scientific literature. Among these, the signature verification strategy has been proposed as a viable approach to prevent CPA by verifying the genuineness of content using digital signatures [16,17]. In this connection, authors in [11] proposed Public Key Digests (PPKD) information of the content publisher. This approach uses an interest packet self-authentication verification system. In response to a packet, the intermediate node verifies the PPKD of the interest packet and identifies the CPA. However, verifying each packet at the intermediate node VNDN is infeasible due to the vehicles' mobility and intermittent connectivity. Other strategies include popularity-based content caching [18,19], reputation-based systems [20], and trust management schemes [21]. An initial attempt to address these challenges was made in [20], which proposed an in-network trusted content caching system based on a reputation system. Although this approach is essential, it does not adhere to the native content exchange mechanism of NDN. At present, less effective or partial solutions to detect and prevent CPA have been considered in the literature. Moreover, Machine-Learning (ML)-based prediction systems have been exploited in various areas, including e-commerce [22], fraud detection, disease diagnosis [23] and so forth. However, no research contribution has yet been made to collectively cope with ML-based security attacks, privacy, routing, and caching problems in VNDN.

Unlike traditional rules-based techniques in literature, we propose a novel and efficient approach to identify and prevent CPA attackers using highly accurate Machine Learning (ML) algorithms. In this research, we initially enable the Roadside Units (RSUs) to collect and store every host vehicle's reputation (positive or negative) in a blockchain database. Subsequently, RSUs perform ML algorithms over the aggregate reputation of host vehicles and classify them as attackers or non-attacker. In order to perform ML classification, there are various publicly available datasets [24]. The most recent real-time dataset for V2V is proposed in [25], including car tracking, cooperative perception tasks, and localization. Another recent work presented a dataset for accident prediction [26]. In order to detect CPA, we employed a publicly accessible BurST-Australian dataset for Misbehavior Detection (BurST-ADMA) [27]. Based on classification results, we provide a content caching algorithm implemented at the intermediate nodes that queries the legitimacy of host vehicles from RSU and determines to store or discard the content. Moreover, we propose a content dissemination system that uses the pub-sub mechanism to disseminate host vehicles' reputations among RSUs.

Thus, the motivation of this study is to detect and thwart attacker attacker vehicles from serving malicious content and ensure the trustworthiness between vehicles. This research aims to enhance the privacy of every vehicle and propose a reliable content dissemination and caching solution. For this paper, the main contributions are as follows:

(1) We propose an ML-based CPA identification and mitigation mechanism that enables the intermediate nodes to store or drop the packet based on the content provider legitimacy.
(2) We use five (5) different ML classifiers and compare their accuracy in identifying attacker and non-attacker vehicles.
(3) We incorporate blockchain to ensure the privacy of vehicles. Our blockchain system maintains the credibility of each vehicle in distributed ledger.
(4) We enhance default NDN's scope from pull to a push-based content dissemination using pub-sub mechanism.

The remainder of this article is organized as follows: Section 2 outlines the existing work. In Section 3, we propose a content caching architecture. Section 4 presents our ML framework for reputation evaluation. Section 5 exhibits experimental results. Lastly, Section 6 ends the article with a conclusion.

2 Related Work

Over the years, a wide range of literature has focused on detecting anomalies and attacks in VNDN [28]. The existing research witnesses a significant contribution in detecting and preventing CPA. In particular, efficient content caching in VNDN [29] has been explored in the literature. The strategies for effective CPA detection includes popularity-oriented content caching [30], cooperative caching [31] and rating-based trust management systems [32,33]. Conversely, ML has been incorporated into VANET for CPA detection and prevention. Hence, we divided the related work into the non-learning rule-based approach and the learning-based system for CPA detection and prevention.

2.1 Rule-Based CPA Mitigation

The non-learning-based CPA detection mechanism involves a static and pre-defined threshold-based rule. A content producer is considered trusted in this architecture if it meets a user-defined threshold value. In a non-learning-based attack detection system, the intermediate nodes determine the reputation value of host vehicles before caching content. To this end, authors in [34] proposed a rating-based blockchain scheme to detect and prevent CPA in NDN. This scheme assigns the ratings (positive or native) to the node based on their legitimacy. This work proposed a honeyguide search algorithm that is derived from a biological rule-based system. This mechanism assigns an initial reputation to each node that is updated as per behavior of the content-producing node. This work includes a Malicious Vehicle Table (MVT) that contains a list of CPA. The intermediate nodes query the reputation of CPA from MVT before caching content. Rezaeifar et al. [33] introduced a CPA detection scheme using trust management in NDN. This mechanism identifies invalid content by defining three metrics: the node's credibility, the content's popularity, and the negative feedback the content consumers provide. Similarly, Khelifi et al. [20] presented a caching scheme based on reputation in VNDN that utilizes blockchain technology to enable secure and efficient caching. This scheme involves an intermediate node caching content if the provider meets a threshold reputation value. Subsequently, the authors extended the scope of their current work in [35] by integrating IFA. The reputation-based content caching policy in both references is based on an algorithm that decides whether to allow the data packet to be cached and forwarded back to the consumer or to drop the packet. Ghali et al. [36] introduced a ranking-based content caching scheme that assigns the ranking to cached content based on an exclusion field provided by consumers. This field includes three metrics: the number of exclusions, the exclusion time distribution, and the exclusion ratio. By evaluating these

metrics, a router determines the ranking of stored content and provides high-ranked content to the consumer.

In addition, reference [18] introduced a Most Popular Content (MPC) system, which enables the nodes to store the most popular content. The content popularity depends on the hit ratio at each node. In order to achieve this goal, each node stores a local popularity table containing the name of content and its corresponding popularity score. If a piece of content exceeds a pre-defined popularity threshold, the corresponding node will cache it in its local cache. However, this scheme is vulnerable to content pollution attacks, where adversary nodes can repeatedly send interests for unpopular content, which leads the nodes to cache unwanted content. In [37], authors proposed trust-based traffic validation information, where RSUs collect traffic events from vehicles and evaluate the legitimacy of the information by broadcasting the same information among vehicles passing through the corresponding area. Once the collected data is identified as valid, the host vehicle will be considered reputed and notified about the event to all the adjacent vehicles. The authors integrated blockchain with VANET [38] to securely store the reputation scores of each vehicle. In this scheme, all the vehicles collectively elect a temporary miner based on the threshold value defined by Trusted Authority (TA). The elected miner has the right to broadcast the block, followed by packing the transactions, which are then verified by other receiving vehicles before being stored in the blockchain. Sun et al. [39] presented a trust-based scheme for evaluating the reputation of vehicles. They used Kalman filtering and chi-squared tests to identify the trusted vehicles in VANET. Based on the reputation of a vehicle, the neighboring vehicles determine whether to accept or reject the information. Similarly, Yang et al. [40] proposed a reputation system for determining the content's reputation in VANET. In this work, ratings are handed over to the content providers and further disseminated to a provisionally chosen node as blocks. Thus, these nodes disseminate the reputation among other nodes. Despite significant contributions, the works mentioned above have limitations, as mentioned in Table 1 below:

Table 1: Rule-based CPA detection

Reference	Environment	Technique	Limitations
[34]	VNDN	Rating-based CPA detection	The rule-oriented CPA detection mechanism is considered in these works. The threshold-based static pre-determined system for assessing the vehicles' behavior is determined. The content producers in these architectures are trusted if they meet pre-configured static reputation scores. However, this mechanism is infeasible in the dynamic network where reputation scores change as time elapses. Moreover, these architectures lack granularity, which may lead to unfairness and discourage vehicles from participating in such a system
[33]	NDN	Credibility and popularity	
[20]	VNDN	Threshold based reputation system	
[35]	VNDN	Blockchain-based CPA and IFA detection	
[36]	NDN	Ranking-based IFA prevention	
[18]	NDN	Popularity-based content caching	
[37]	VANET	Trust-based content sharing	
[38]	VANET	Reputation-based	
[40]	VNDN	Threshold-based data sharing	
[39]	VANET	Trust-based con	

2.2 Learning-Based Attack Mitigation

ML and Deep Learning (DL) have rapidly developed in different research areas, such as attack detection, image inpainting models [41] and fraud detection. Specifically, ML is most accurate for classification problems, whereas DL has obvious advantages in complex problems such as image

[42] and voice recognition. The learning-based attack detection mechanism involves ML algorithms for classifying attackers and non-attackers with high accuracy. Several ML-based attack detection mechanisms proposed in the literature, such as authors in [43], evaluated the performance of Logistic Regression (LR) and Support Vector Machine (SVM) for misbehavior detection. Their results exhibit that SVM performs better than LR in terms of accuracy. Similar to our research, reference [44] exploits a reputation-based misbehavior detection system, where every vehicles' reputation is shared with RSUs. The RSUs perform ML classification to validate the host vehicle's reputation. When a vehicle is classified as an attacker, the RSU propagates such information among all the neighboring RSUs and vehicles. Similarly, the authors in [45] employed binary and multi-classification techniques for detecting and identifying attack types. Their significant contributions showed promising results for detecting misbehaving vehicles in traditional VANETs. In [27], authors proposed a BrustADMA dataset and evaluated the performance of different classifiers. Based on the results, RF performed excellently compared to other ML algorithms. Table 2 illustrates the summarized ML-based related work and their limitations in attack detection systems.

Table 2: ML-based CPA detection

Reference	Environment	Limitations
[43]	VANET	The authors evaluated the SVM and RL classifiers only. This contribution lacks ML classifiers.
[44]	VANET	This approach uses four ML classifiers. However, no CPA detection in NDN is exploited.
[46]	VANET	This work evaluated KNN and RF classifiers only. The proposed classifiers' accuracy is not highly accurate.
[45]	VANET	Although the authors achieved high accuracy, they did not exploit ML in VNDN, which is different from traditional VANET.
[27]	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.

Compared to existing contributions, which focus on an individual issue or have not adequately addressed the ML-based attacks detection solutions in VNDN, our proposed approach addresses multiple challenges by leveraging ML to detect and predict vehicles' reputations and integrating blockchain technology to store this information in a VNDN environment securely. Our method mitigates CPA by allowing intermediate nodes to store or neglect the data packet based on prior legitimacy detected by ML classifiers. See Table 3 for a list of notations used in our research.

Table 3: Summary of notations

Notation	Description
$I_P$kt	Interest packet
$D_P$kt	Data packet
C_P	Content producer
C_C	Content consumer

(Continued)

Table 3 (continued)

Notation	Description
CP_R	Content producer reputation
T_C	Trusted content
CP_R^n	Content producer new reputation
$CP_R^n - 1$	Content producer previous reputation
$AgrCP_R$	Aggregate content producer reputation
T_X	Transaction

3 Proposed System Model

This section is mainly divided into three parts. First, we introduce the primary components and their respective responsibility in our proposed scheme. Afterward, we propose a content forwarding and caching mechanism based on Algorithm 1, aiming to detect and prevent CPA In addition, we present Algorithm 2, which is designed for content validation and dissemination among RSU through the utilization of a suitable naming structure. Finally, we propose a blockchain oriented block dissemination and validation scheme.

3.1 System Components

Our proposed mechanism mainly comprises three entities: the TA, vehicles, and RSUs. Each entity plays an important role in ensuring content transmission in VNDN.

TA: In our proposed mechanism, the TA is the governmental traffic department responsible for registering and authenticating vehicles and RSUs. The TA generates a pair of public and private keys for each registered entity, which are used to encrypt and decrypt blockchain transactions.

Vehicles: The OBU-equipped vehicles are the main entities in our proposed mechanism. The vehicle can be content consumers, intermediate nodes or content producers. The vehicles are further divided into two categories: public buses and cars. Depending on the situation, the role of vehicles in our proposed model is interchangeable. For example, a content consumer can be a relay node or producer in another situation.

RSU: The role of RSU is pivotal in our proposed system; they serve as important infrastructure elements that help to support communication between vehicles and facilitate the delivery of services to vehicles. RUSs are equipped with superior computational power and uninterrupted stability in the network. All RSUs have been designated to gather and anticipate reputation data through ML algorithms. We assume all the RSUs are blockchain nodes.

3.2 Content Forwarding and Caching Scheme

As indicated in Fig. 2, the content consumer initiates the communication process by broadcasting an interest packet to request specific content. When a vehicle receives an interest packet, it first verifies its cache to see if it has the requested content. If the content is in its CS, it will immediately send a data packet containing the content back to the requester. If the content is not in the CS, the interest packet will be forwarded further into the network until it reaches a producer. A vehicle with the matching content transmits a data packet to the intermediate vehicle, which contains additional information, such as the producer's cryptographic signature, identification details, and other optional fields. The

intermediate vehicle then queries its local storage to check the aggregate reputation of the content producer. It is to be noted that every vehicle can obtain an updated reputation of vehicles from the nearest RSU at any time. If the reputation of the content producer is found to be an attacker, the intermediate vehicle immediately discards the data packet without caching and forwarding it to the content consumer. The RSU utilizes ML to detect the reputation and classify the vehicles as either attacker (0) or legitimate (1). Based on this information, the intermediate vehicle determines whether to store and provide the content to the consumer or discard the packet.

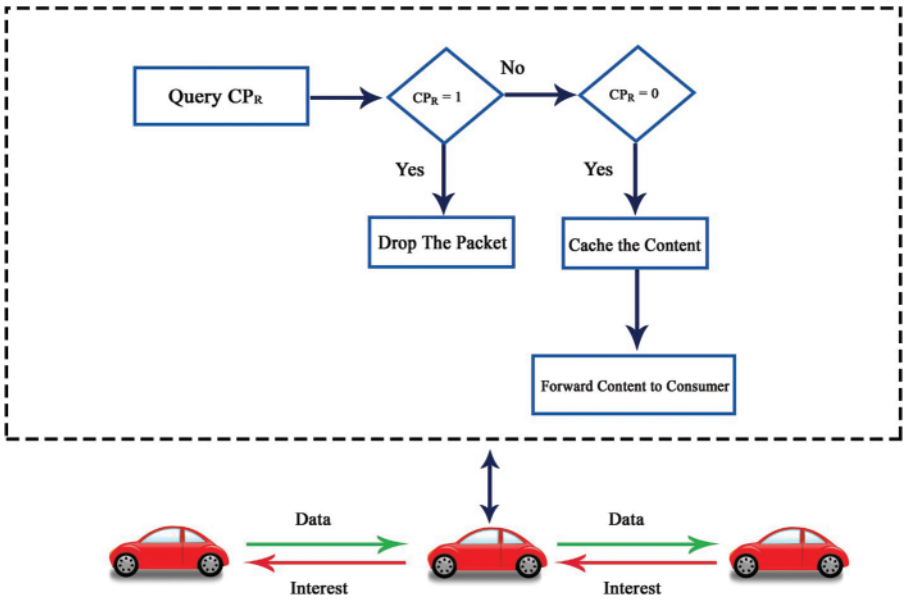

Figure 2: Content caching mechanism

Our proposed Algorithm 1 verifies the legitimacy of host vehicle before storing content at the intermediate node. The intermediate vehicle caches the content in its CS and transmits it to the content consumer only when the reputation value obtained from the RSU is "Legitimate" (0). Conversely, the intermediate vehicle discards the data packet if the reputation value received is "Attacker" (1).

Algorithm 1: Content caching policy at intermediate node

Require: $D_P$kt from C_P
Query aggregate reputation
if $CP_R = 1$ then
    Drop the packet
else if $CP_R = 0$ then
    Cache content
    Forward D_P to C_C
end if

3.3 Content Validation Policy

Following Algorithm 2, once the content consumer vehicle receives the data packet, it verifies the content's legitimacy. If the content is deemed legitimate, a non-attacker identification (0) is rewarded

to the legitimate vehicle and communicated to the RSUs. However, if the content is identified as malicious, an attacker classification (1) is charged to the producer and shared with the RSUs.

Algorithm 2: Content validation policy

```
Require: D_Pkt
if D_Pkt = T_C then
        CP_R^n ← CP_R^(n−1) + 0
        Push CP_R^n to pre-subscribed RSUs
else if If D_Pkt ≠ T_C then
        CP_R^n ← CP_R^(n−1) + 1
        Push CP_R^n to pre-subscribed RSUs
end if
```

3.4 Reputation Dissemination

The consumer leverages pub-sub and propagates the content producer's reputation to the pre-subscribed RSUs. In our proposed system model, a pub-sub enables the NDN nodes to broadcast the content among pre-subscriber nodes without expecting an interest packet for every content. The proposed system allows content consumer vehicles to propagate reputation (either 1 or 0) to the nearest RSUs. Due to the dynamic and uncertain reputation generation time, RSUs cannot send a subscription request for each reputation individually. Determining too short or frequent broadcast time will deteriorate the network performance. In order to address this problem, an appropriate naming structure is outlined below.

3.5 Naming Structure

NDN uses opaque naming structure that allows users to construct their naming hierarchy as needed. Leveraging opaque naming structure of NDN, we designed naming for content dissemination, reputation dissemination, and subscription. To avoid sending multiple subscription requests for each reputation, the RSUs subscribe to all versions of reputations using a prefix naming structure. We propose the following naming structures as per requirement of content and reputation dissemination.

Reputation retrieval request: VNDN/Reputation/Vehicle/No/MustBeFresh

Content retrieval request: VNDN/Infotainment/Music/Album/XYZ.mp4

Reputation subscription request: VNDN/Pub-Sub/Reputation

The reputation retrieval request with the "MustBeFresh" keyword indicates that a vehicle requires a reputation for a specific vehicle with the most recent reputation. Similarly, the naming structure in content retrieval requests depicts a node requesting music named "XYZ" in the infotainment section. On the other hand, the reputation subscription request is a prefix that informs the vehicles to deliver the reputation of every vehicle. The prefix is a broad naming structure that allows vehicles to send the reputation of every vehicle.

3.6 Block Dissemination

As mentioned earlier, the primary responsibility of RSUs is to store and compute the vehicles' reputation within the blockchain network Once RSU receives reputation, it aggregates and classifies the reputation using ML algorithms. Once the reputation is determined, the RSU creates a transaction and adds it to a new block in the blockchain. Consequently, the newly generated block is disseminated

among all nodes within blockchain network that ensures the a secure reputation in a decentralized manner. Afterward, each blockchain nodeverifies the by exploiting a consensus algorithm. In order to varify blocks, we consider a Proof of Work (PoW) consensus mechanism, a most secure and renowned algorithm, where nodes are required to mine the block with complex calculations.

4 Reputation Evaluation Framework

This section assesses the performance of ML classifier in detecting the credibility of vehicles by using a publicly available dataset at RSU. The analysis mains to employ ML classification techniques to identify attackers and legitimate vehicles using binary classification on the dataset. The motive of our proposed research is to identify misbehavior of vehicles by using ML classifiers and performance metrics, including precision, recall, and F1 score, to assess the accuracy of the models. The ultimate motive of this research is establish an accurate reputation detection and prevention framework to enhance VANET's security and efficiency. Moreover, our proposed reputation evaluation identifies the best ML classifier for accurately detecting attacks.

4.1 Method and Data Source

As illustrated in Fig. 3, our proposed system investigates the reputation of vehicles, wherein RSUs are exploited for collecting and storing every individual host vehicles' reputation. The collected reputation scores are then used to train and test various ML classifiers to distinguish between benign and attacker vehicles. In order to achieve this, the reputation evaluation process is divided into three phases: dataset collection, data preparation, and performance evaluation of classifiers.

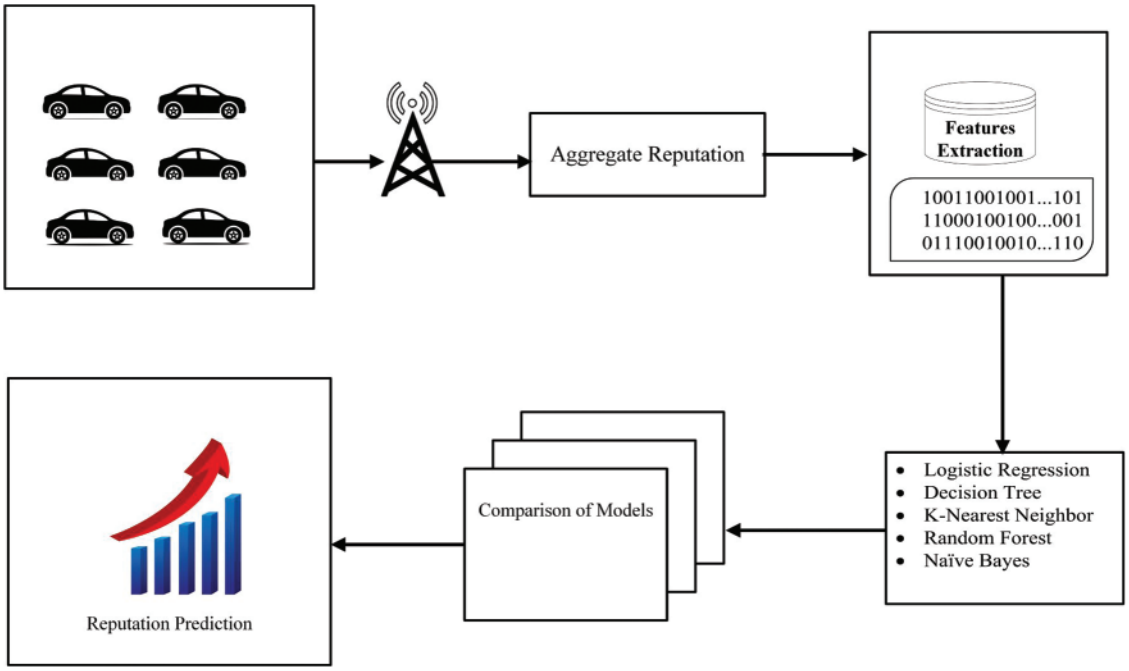

Figure 3: Proposed framework model

4.2 Dataset Collection

In order to identify the behavior of vehicles, we utilized the publicly available BurST-Australian dataset for Misbehavior Analysis (BurST-ADMA) [27]. This dataset was simulated using the Urban Mobility (SUMO) in the Burwood suburbs of Melbourne, Victoria, Australia. The dataset comprises seven entities: bicycles, buses, motorcycles, pedestrians, public transport buses and trams, trucks, and passenger vehicles. It contains 207,315 messages, of which 179,126 are legitimate and 28,189 are malicious. Although BurST-ADMA contains seven different types of attacks, our research is focused on identifying malicious vehicles, irrespective of the attack type. Therefore, we treat all the attack types as a single attack, represented by (1), while legitimate entities are represented by (0). The dataset provides seven features, including timestamp, vehicle ID, X and Y coordinates, speed, heading, and attack type. We used all features in our proposed research.

4.3 Data Preprocessing

In this stage, we consolidated the seven different attack types (1–7) into one attack type. We categorized them into binary labels of either "attacker" (1) or "legitimate" (0) for classification. To ensure the accuracy of our proposed model, we randomly divided the dataset into 70% (145,120 messages) for training and 30% (53,737 messages) for testing and employed 10-fold cross-validation to assess the model's performance. We first partitioned the entire dataset into k-folds and then utilized 10-fold cross-validation to estimate the performance of our proposed algorithm. This method enables to demonstrate the accuracy of our proposed ML-based approach for accurately detecting and preventing the reputation of host vehicles.

4.4 Classification

It is supervised ML that is used to allocate a known input data point to one of the pre-determined classes or categories. The classification method is initially trainedon a preprocessed dataset where each data point is labeled with its corresponding class. The model learns from the underlying patterns and relationships between the input features and the class labels during the training phase and then uses the learned knowledge to classify the new, unseen data instances known as testing. In order to identify and anticipate vehicles' actions, we use the following ML classifiers.

4.4.1 LR Classification

The LR [47] classification model utilized to predict the probability of categorical data. LR uses a logistic function for calculating the probability for two class classification. The expression for logistic function in binary classification is represented as:

$$\rho\,(\mathrm{X}) = \frac{e^{\alpha+\beta \mathrm{X}}}{1 + e^{\alpha+\beta \mathrm{X}}} \tag{1}$$

where α and β are linear prediction, X is an independent variable, and α and β are linear predictors.

4.4.2 Decision Tree (DT) Classification

The DT is a well-known ML classifier, which categorizes samples based on their feature values. The DT creation involves investigating training samples and deciding the best features to separate the data into subsets based on a specific principle, such as information gain or the Gini index. The motive is to establish a tree that can accurately calculate the results of latest data based on the existing features. The DT begins with a root node or vertex, encompassing features selected using attribute

selection metrics like the Gini Index and Information Gain [48]. The Gini Index is expressed as:

$$Gini\,(D) = 1 - \sum_{n-1}^{m} P_i^2 \tag{2}$$

where P_i is a probability that a feature vector belonging to an attacker.

4.4.3 K-Nearest Neighbor (KNN) Classification

The KNN classifier [49] is a renowned ML algorithm for managing large datasets. It is a straightforward and flexible ML classifier that can be utilized for both classification and regression tasks. It is able to operate multi-class classification and straightforwardness makes it a good choice for different applications. KNN groups latest data points in the training set, delegates the latest data point to the most common class among those K neighbors, and returns the majority or average label of those neighbors. The best value of K varies based on the specifics of the dataset.

4.4.4 Random Forest (RF)

The RF [50] ML classifier is one of the most powerful ML classifier that combines different DTs' outputs using majority voting. This method results in a more robust solution to difficult problems, and the prediction is made by taking the mean of the outputs from individual DTs.

4.4.5 Gaussian Naive Bayes (GNB) Classification

The GNB algorithm is a straightforward classification method that uses Bayes' theorem to predict the class of unlabeled data points. It calculates the prior probabilities of the classes and applies them to new, unseen data. The independence of the features in GNB makes it a simple and computationally efficient approach.

4.5 Model Evaluation

We conducted an experimental evaluation on the BurST-ADMA dataset to evaluate the effectiveness of various ML classifiers. We evaluated the performance of these classifiers using accuracy, precision, recall, and the F1 score as evaluation metrics. The corresponding mathematical models for each algorithm are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{6}$$

The evaluation consists of four performance metrics: True positive (TP), which represents the correctly classified 'Positive Reputation' of positive samples; False positive (FP) is classified as a proportion of samples if they are not in the class; True negative (TN) is the correctly classified as a negative proportion of samples, while False negatives (FN) is the samples which are as classified as a positive proportion of samples.

5 Experimental Results and Discussion

Our research aims to identify and prevent CPA in VNDN. Thus, we accomplished this by evaluating the performance of various ML algorithms considering the vehicles' behavior. Table 4 illustrates our proposed model's precision, recall, and F1 score. In order to assess the performance of each ML algorithm, we utilized a precision-recall curve in conjunction with a Receiving Operative Characteristics (ROC) curve to visualize the obtained outcomes. The precision and recall curves are mainly used for evaluating binary classification performance. The curve in ROC indicates the trade-off between precision and recall values, with a larger area under the curve suggesting high 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 illustrated in Table 4, RF, DT and KNN yielded outstanding accuracy in detecting CPA attackers. On the other hand, LG and GNB also showed satisfactory results.

Table 4: Performance evolution of various ML algorithms

ML classifiers	Precision	Recall	F1 score
LR	0.82	0.78	0.80
DT	1.00	0.99	1.00
KNN	1.00	0.97	0.98
RF	1.00	1.00	1.00
GNB	0.54	1.00	0.70

5.1 Visualized Results

To gain insights into the performance of our ML models, we visualized our results using accuracy calculation techniques with a ROC. First, we employed precision-recall and ROC curves to assess the trade-off between precision and recall for binary classification. The Area Under Curve (AUC) indicated strong performance in three models, with the highest AUC achieved by the RF, DT, and KNN models. In contrast, GNB and LG depicted average performance. According to the experimental findings, ML classifiers such as RF, DT, and KNN exhibited outstanding performance in classifying the attacker and legitimate vehicles with high accuracy. Figs. 4 to 6 show the precision-recall curve performances for DT, KNN, and RF classifiers, respectively. The precision-recall in these classifiers shows a perfect curve, reflecting the highest accuracy. LR in Fig. 7 and GNB in Fig. 8 illustrate a satisfactory performance. Fig. 9 depicts the consolidated and comparative evaluation.

5.2 Comparative Performance Evaluation

Existing literature uses different ML techniques to detect and prevent attacks and ensure security and privacy in VANET. However, none of them evaluated ML classifiers for attack detection in VNDN. Thus, our research has a novel contribution to detecting and preventing CPA in VNDN. Our comparative performance analysis aims to evaluate several ML classifiers for attack detection and compare them with existing related work in VANET. Our comparative analysis in Table 5 shows that our proposed RF classifier outperformed all other related works, achieving 100% accuracy in detecting attacks. These findings have important implications for the development of effective attack detection systems. The proposed RF classifier has the potential to significantly enhance the accuracy and reliability of such systems, thereby improving their effectiveness in detecting and mitigating

attacks. Overall, our proposed ML-based attack detection and prevention has a novel contribution to preventing attacker vehicles from serving malicious content.

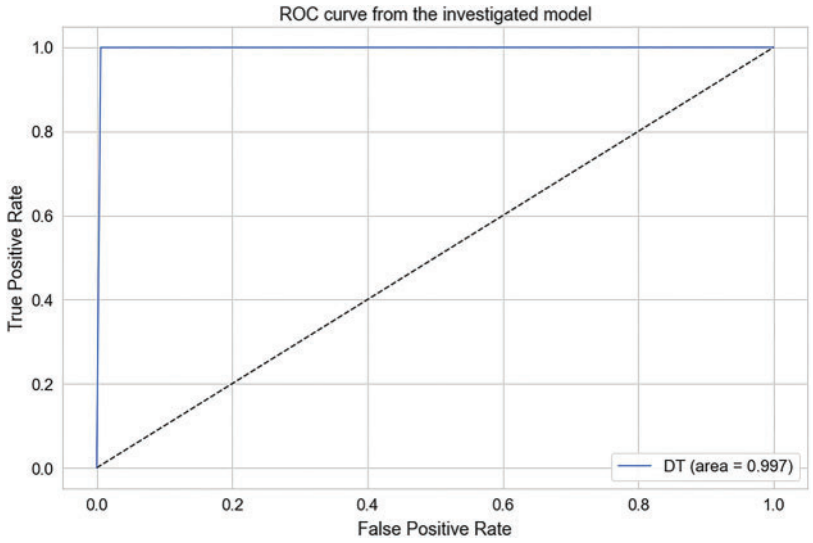

Figure 4: DT accuracy score

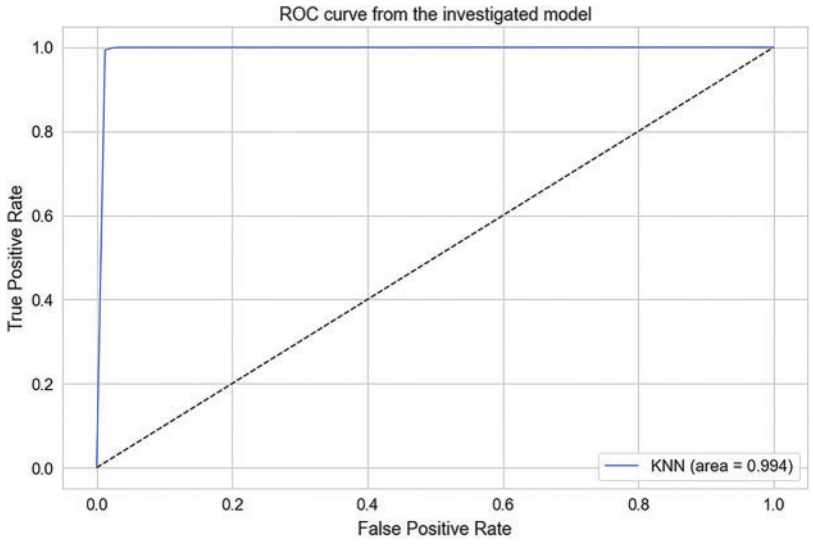

Figure 5: KNN accuracy score

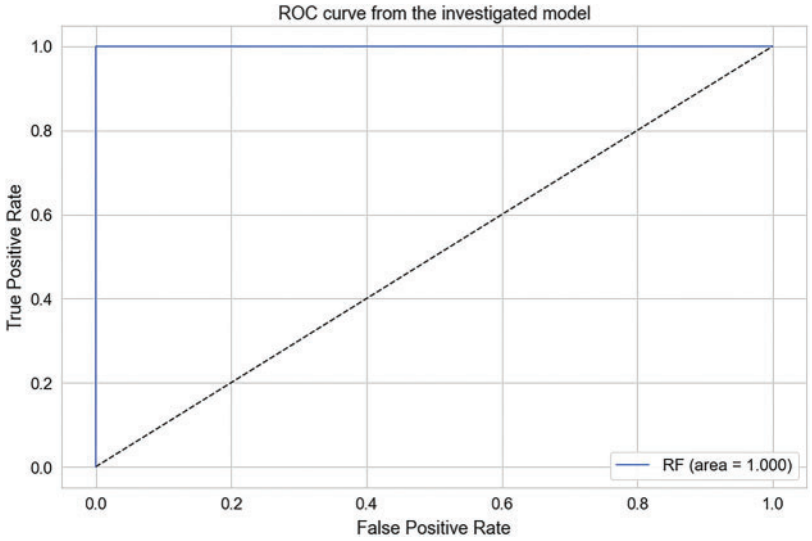

Figure 6: RF accuracy score

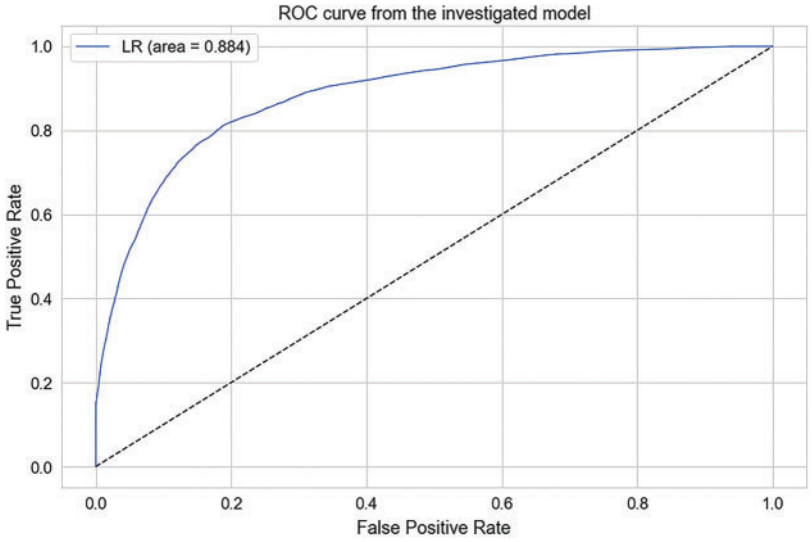

Figure 7: LR accuracy score

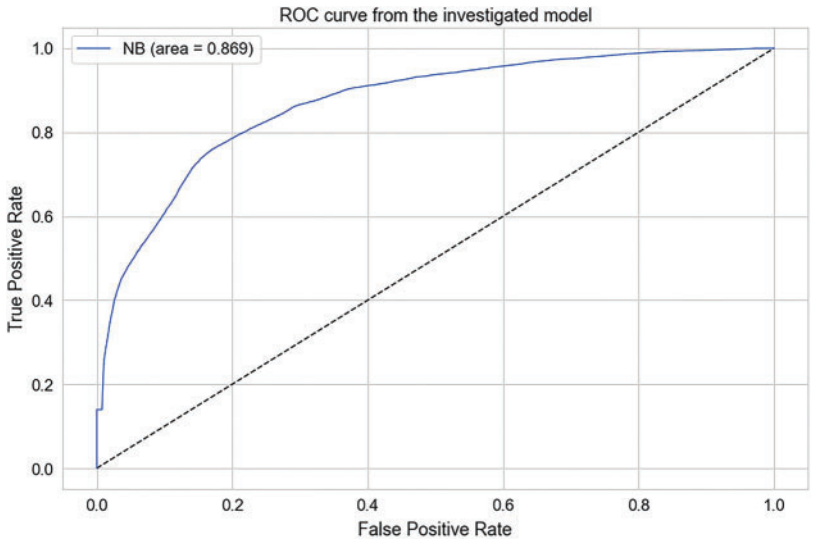

Figure 8: GNB accuracy score

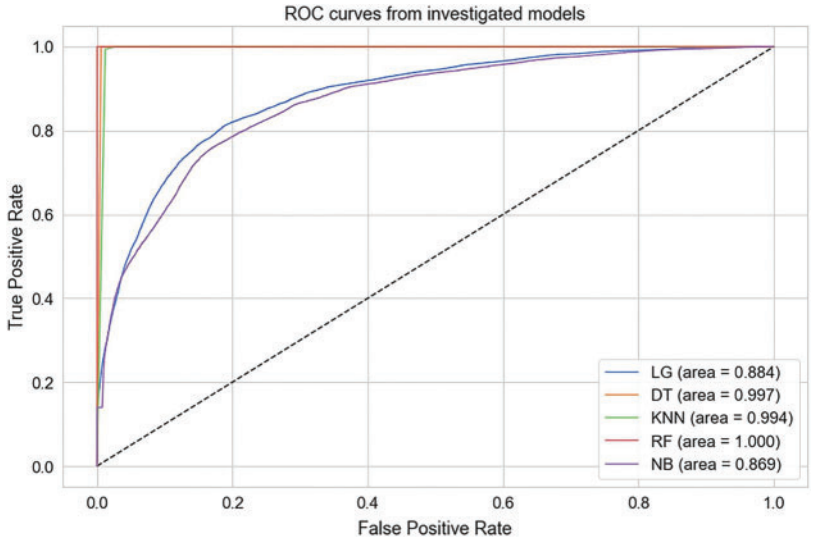

Figure 9: Consolidated accuracy scores

Table 5: Performance evolution of various related works

References	Dataset	Environment	Best ML classifier	Accuracy
[43]	VeReMi	VANET	SVM	97%
[51]	BurST-ADMA	VANET	RF	99.6%
[52]	VeReMi	VANET	RF	90.83%
[53]	Knowledge discovery in databases (KDD)	VANET	RF	93.77%

(Continued)

Table 5 (continued)

References	Dataset	Environment	Best ML classifier	Accuracy
[54]	BurST-ADMA	VANET	Optimizable AdaBoost	98.93%
[55]	BurST-ADMA	VANET	RF	99.53%
[56]	Real-time	VANET	RF	99%
Proposed	BurST-ADMA	VNDN	RF	100%

6 Conclusion

Security and privacy are essential research areas in VNDN as they directly involve human lives. In order to protect sensitive information from unauthorized access, a secure network architecture must be implemented to ensure the safety and confidentiality of passengers and drivers. Specifically, CPA attacks have dire consequences, including road accidents. In order to tackle CPA in VNDN, several rule-based research contributions are presented in the literature. However, none of them exploited ML features to classify and prevent CPA in VNDN. In this paper, we mainly focused on identifying and preventing CPA attacks using the potential of ML classifiers. Based on our proposed ML classification results, our reputation-based content caching algorithm allows or prevents vehicles from serving content. In addition, our proposed system allows vehicles to propagate the reputation information of host vehicles using push-based content dissemination using a pub-sub approach among blockchain-empowered RSUs. As a result, RSUs collect reputations in binary form and perform an ML algorithm on the aggregate reputation. Upon querying reputation, the RSUs provide classification results to the intermediate nodes. We implement our proposed ML algorithm on the publicly available dataset BurST-ADMA. Our experimental findings demonstrate that RF has achieved 100% accuracy in detecting attackers. Meanwhile, DT achieved 99.7%, KNN obtained 99.4% accuracy, GNB achieved 86.9%, and LR provided 88.4% accuracy. Hence, our proposed research identifies and prevents attacker vehicles, ensures the privacy of vehicles, and enhances NDN's scope from pull to push-based content dissemination using the pub-sub approach. In addition, this research is limited to ML classification over the simulation-based dataset. However, this approach can be further implemented in real-time scenarios for reliable and trusted content dissemination between vehicles in the future.

Acknowledgement: The authors acknowledge the Researchers Supporting Project Number (RSPD2023R553), King Saud University, Riyadh, Saudi Arabia.

Funding Statement: This research is funded by Research Supporting Project Number (RSPD2023R553), King Saud University, Riyadh, Saudi Arabia.

Author Contributions: Study conception and design: A. H. Magsi, A. Ghulam, S. Memon; data collection: A. H. Magsi, K. Javeed, M. Alhussein, I. Rida; analysis and interpretation of results: A. Ghulam, S. Memon, I. Rida; draft manuscript preparation: A. H. Magsi, A. Ghulam, K. Javeed, M. Alhussein; data curation: A. H. Magsi, S. Memon, I. Rida; visualization: A. H. Magsi, A. Ghulam, M. Alhussein; resources: A. H. Magsi, S. Memon, K. javed, M. Alhussein. 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 first author 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 & Tutorials*, vol. 18, no. 1, pp. 263–284, 2015.

[2] K. B. Kelarestaghi, M. Foruhandeh, K. Heaslip and R. Gerdes, "Survey on vehicular ad hoc networks and its access technologies security vulnerabilities and countermeasures," arXiv preprint arXiv:1903.01541, 2019.

[3] Y. Y. Nasrallah, I. Al-Anbagi and H. T. Mouftah, "Enhanced algorithms for the IEEE 802.11 p deployment in vehicular ad hoc networks," in *Proc. of VTC-Fall*, Toronto, ON, Canada, pp. 1–5, 2017.

[4] L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, K. Claffy *et al.,* "Named data networking," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 3, pp. 66–73, 2014.

[5] G. Xylomenos, C. N. Ververidis, V. A. Siris, N. Fotiou, C. Tsilopoulos *et al.,* "A survey of information-centric networking research," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 2, pp. 1024–1049, 2013.

[6] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs *et al.,* "Networking named content," in *Proc. of Co-NEXT*, Rome, Italy, pp. 1–12, 2009.

[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 & Tutorials*, vol. 22, no. 1, pp. 320–351, 2020.

[8] M. Amadeo, C. Campolo, A. Iera and A. Molinaro, "Named data networking for IoT: An architectural perspective," in *Proc. of EuCNC*, Bologna, Italy, pp. 1–5, 2014.

[9] M. Amadeo, C. Campolo, A. Molinaro and N. Mitton, "Named data networking: A natural design for data collection in wireless sensor networks," in *Proc. of 2013 IFIP WD*, Valencia, Spain, pp. 1–6, 2013.

[10] C. A. Kerrche, F. Ahmad, M. Elhoseny, A. Adnane, Z. Ahmad *et al.,* "Internet of vehicles over named data networking: Current status and future challenges," in *Emerging Technologies for Connected Internet of Vehicles and Intelligent Transportation System Networks*. Cham, Switzerland: Springer, pp. 83–99, 2020.

[11] K. Lei, J. Fang, Q. Zhang, J. Lou, M. Du *et al.,* "Blockchain-based cache poisoning security protection and privacy aware access control in ndn vehicular edge computing networks," *Journal of Grid Computing*, vol. 18, pp. 593–613, 2020.

[12] A. Benmoussa, C. A. Kerrache, N. Lagraa, S. Mastorakis, A. Lakas *et al.,* "Interest flooding attack in named data networking: A survey of existing solutions, open issues, requirements, and future directions," *ACM Computing Surveys*, vol. 55, no. 7, pp. 1–37, 2022.

[13] M. Avijit, "Man-in-the-middle-attack: Understanding in simple words," *Cyberspace : Jurnal Pendidikan Teknologi Informasi*, vol. 2, pp. 109–134, 2019.

[14] M. C. Bidóia, M. A. Cavenaghi, R. Spolon, R. Spolon, A. Manacero Jr *et al.,* "Simulation of a centralized reputation system for vanets," in *Proc. of PDPTA*, Las Vegas, USA, pp. 1, 2014.

[15] I. A. Kapetanidou, C. A. Sarros and V. Tsaoussidis, "Reputation-based trust approaches in named data networking," *Future Internet*, vol. 11, no. 11, pp. 241, 2019.

[16] Y. Wang, Z. Qi, K. Lei, B. Liu and C. Tian, "Preventing "bad" content dispersal in named data networking," in *Proc. of ACM-TUR-C*, Shanghai, China, pp. 1–8, 2017.

[17] Y. Yu, Y. Li, X. Du, R. Chen and B. Yang, "Content protection in named data networking: Challenges and potential solutions," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 82–87, 2018.

[18] C. Bernardini, T. Silverston and O. Festor, "MPC: Popularity-based caching strategy for content centric networks," in *2013 IEEE ICC*, Budapest, Hungary, pp. 3619–3623, 2013.

[19] J. Li, H. Wu, B. Liu, J. Lu, Y. Wang *et al.,* "Popularity-driven coordinated caching in named data networking," in *Proc. of ACM/IEEE ANCS*, Texas, USA, pp. 15–26, 2012.
[20] H. Khelifi, S. Luo, B. Nour, H. Moungla and S. H. Ahmed, "Reputation-based blockchain for secure ndn caching in vehicular networks," in *Proc. of IEEE CSCN*, Paris, France, pp. 1–6, 2018.
[21] G. Rathee, A. Sharma, R. Kumar, F. Ahmad and R. Iqbal, "A trust management scheme to secure mobile information centric networks," *Computer Communications*, vol. 151, pp. 66–75, 2020.
[22] K. Shaukat, S. Luo, N. Abbas, T. M. Alam, M. E. Tahir *et al.,* "An analysis of blessed friday sale at a retail store using classification models," in *Proc. of ICSIM*, Yokohama, Japan, pp. 193–198, 2021.
[23] T. M. Alam, K. Shaukat, A. Khelifi, H. Aljuaid, M. Shafqat *et al.,* "A fuzzy interface-based decision support system for disease diagnosis," *The Computer Journal*, 2022. [Online]. Available: https://academic.oup.com/comjnl/advance-article-abstract/doi/10.1093/comjnl/bxac068/6603446
[24] T. M. Alam, K. Shaukat, I. A. Hameed, S. Luo, M. U. Sarwar *et al.,* "An investigation of credit card default prediction in the imbalanced datasets," *IEEE Access*, vol. 8, pp. 201173–201198, 2020.
[25] R. Xu, X. Xia, J. Li, H. Li, H. Zhang *et al.,* "V2V4Real: A real-world large-scale dataset for vehicle-to-vehicle cooperative perception," arXiv preprint arXiv:2303.07601, 2023.
[26] R. Xu, H. Xiang, Z. Tu, X. Xia, M. H. Yang *et al.,* "V2X-VIT: Vehicle-to-everything cooperative perception with vision transformer," in *Proc. of ECCV*, Tel Aviv, Israil, pp. 107–124, 2022.
[27] 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 *PerCom Workshops*, Pisa, Italy, pp. 624–629, 2022.
[28] O. Slama, B. Alaya, S. Zidi and M. Tarhouni, "Comparative study of misbehavior detection system for classifying misbehaviors on vanets," in *Proc. of CoDIT*, Istanbul, Turkey, vol. 1, pp. 243–248, 2022.
[29] 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 & Tutorials*, vol. 22, no. 4, pp. 2378–2407, 2020.
[30] K. Suksomboon, S. Tarnoi, Y. Ji, M. Koibuchi, K. Fukuda *et al.,* "PopCache: Cache more or less based on content popularity for information-centric networking," in *Proc. of Conf. LCN*, Sydney, NSW, Australia, pp. 236–243, 2013.
[31] L. Yao, Y. Wang, X. Wang and G. W. Wu, "Cooperative caching in vehicular content centric network based on social attributes and mobility," *IEEE Transactions on Mobile Computing*, vol. 20, no. 2, pp. 391–402, 2019.
[32] Z. Yang, K. Yang, L. Lei, K. Zheng and V. C. M. Leung, "Blockchain-based decentralized trust management in vehicular networks," *IEEE Internet Things Journal*, vol. 6, no. 2, pp. 1495–1505, 2019.
[33] Z. Rezaeifar, J. Wang and H. Oh, "A trust-based method for mitigating cache poisoning in named data networking," *Journal of Network and Computer Applications*, vol. 104, pp. 117–132, 2018.
[34] Z. Sabir and A. Amine, "BIoVN: A novel blockchain-based system for securing internet of vehicles over ndn using bioinspired HoneyGuide," in *Advances in Blockchain Technology for Cyber Physical Systems*. Cham, Switzerland: Springer, pp. 177–192, 2022.
[35] 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, pp. 106715, 2020.
[36] C. Ghali, G. Tsudik and E. Uzun, "Needle in a haystack: Mitigating content poisoning in named data networking," in *Proc. of SENT*, New York, USA, pp. 1–10, 2014.
[37] Y. T. Yang, L. D. Chou, C. W. Tseng, F. H. Tseng and C. C. Liu, "Blockchain-based traffic event validation and trust verification for VANETs," *IEEE Access*, vol. 7, pp. 30868–30877, 2019.
[38] M. Li, J. Weng, A. Yang, J. N. Liu and X. Lin, "Toward blockchain-based fair and anonymous ad dissemination in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 11, pp. 11248–11259, 2019.
[39] M. Sun, M. Li and R. Gerdes, "A data trust framework for VANETs enabling false data detection and secure vehicle tracking," in *Proc. of 2017 IEEE Conf. on CNS*, Las Vegas, NV, USA, pp. 1–9, 2017.
[40] Z. Yang, K. Zheng, K. Yang and V. C. M. Leung, "A blockchain-based reputation system for data credibility assessment in vehicular networks," in *Proc. of PIMRC*, Montreal, QC, Canada, pp. 1–5, 2018.

[41] Y. Chen, R. Xia, K. Zou and K. Yang, "FFTI: Image inpainting algorithm via features fusion and two-step inpainting," *Journal of Visual Communication and Image Representation*, vol. 91, pp. 103776, 2023.
[42] Y. Chen, R. Xia, K. Zou and K. Yang, "RNON: Image inpainting via repair network and optimization network," *International Journal of Machine Learning and Cybernetics*, vol. 14, pp. 1–17, 2023.
[43] 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 *ISEA-ISAP*, Jaipur, India, pp. 166–178, 2019.
[44] A. Sharma and A. Jaekal, "Machine learning based misbehavior detection in VANET using consective BSM approach," *IEEE Open Journal of Vehicular Technology*, vol. 3, pp. 1–14, 2021.
[45] 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 Computer Engineering (IJECE)*, vol. 11, no. 3, pp. 2535–2547, 2021.
[46] 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.
[47] C. Y. J. Peng, K. L. Lee and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The Journal of Educational Research*, vol. 96, no. 1, pp. 3–14, 2002.
[48] S. B. Kotsiantis, I. Zaharakis and P. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, no. 1, pp. 3–24, 2007.
[49] Z. Deng, X. Zhu, D. Cheng, M. Zong and S. Zhang, "Efficient knn classification algorithm for big data," *Neurocomputing*, vol. 195, pp. 143–148, 2016.
[50] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[51] G. O. Anyanwu, C. I. Nwakanma, J. M. Lee and D. S. Kim, "Novel hyper-tuned ensemble Random Forest algorithm for detection of false basic safety messages in Internet of Vehicles," *ICT Express*, vol. 9, no. 1, pp. 122–129, 2023.
[52] P. Sharma and L. Hong, "A machine-learning-based data-centric misbehavior detection model for internet of vehicles," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4991–4999, 2020.
[53] M. S. Alsahli, M. M. Almasri, M. Al-akhras, A. I. Al-issa and M. Alawairdhi, "Evaluation of machine learning algorithms for instrusion detection system in WSN," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, pp. 617–626, 2021.
[54] G. O. Anyanwu, C. I. Nwakanma, J. M. Lese and D. S. Kim, "Misbehavior detection in connected vehicles using BurST-ADMA dataset," in *Proc. of ICTC*, Jeju Island, Korea, pp. 874–878, 2022.
[55] G. O. Anyanwu, C. I. Nwakanma, J. M. Lee and D. S. Kim, "Fasification detection system for iov using randomized search optimization ensemble algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4158–4172, 2023.
[56] K. Rashid, Y. Saeed, A. Ali, F. Jamil, R. Alkanhel *et al.,* "An adaptive real-time malicious node detection framework in vehicular ad-hoc networks," *Sensors*, vol. 23, no. 5, pp. 2594, 2023.