
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 6, pp. 4540–4557, DOI 10.12694/scpe.v25i6.3270

ENHANCING BLACK HOLE ATTACK DETECTION IN VANETS: A HYBRID APPROACH INTEGRATING DBSCAN CLUSTERING WITH DECISION TREES

SENG-PHIL HONG*

Abstract. Ensuring the security of communication is crucial in Vehicular Ad Hoc Networks (VANETs) to protect the integrity of information sharing among cars. To implement VANET communication as an answer for the different uses, secure communication is necessary. The unreliability of VANET environments is caused by message delays or tampering in VANET applications. Finding the sweet spot between VANET security and performance and dependability is the primary goal. This project's overarching goal is to fortify VANETs against Blackhole Routing Attacks and, by identifying and blocking harmful nodes, to mitigate the blackhole impact. This paper proposes a robust hybrid approach for the detection of black hole attacks in VANETs, leveraging the synergy between DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering and Decision Trees. DBSCAN, a density-based clustering algorithm, is employed to identify spatial clusters of vehicles, while Decision Trees are utilized to discern normal communication patterns from malicious ones within these clusters. The integration of these two techniques enhances the accuracy and efficiency of black hole attack detection in the dynamic and resource-constrained VANET environment. Experimental results demonstrate the effectiveness of the proposed hybrid approach, providing a promising solution for bolstering the security of VANETs against emerging threats. Here in result 73.89% improvement is received in Packet Drop Rate using DBSCAN, also minor improvement over Throughput and Average end to end delay and major improvement in terms of Network Routing Load.

Key words: VANET security; Black hole attack detection; DBSCAN clustering; Decision Trees; Hybrid approach; Network reliability.

1. Introduction. Moving beyond Mobile Ad Hoc Networks (MANETs), [1] which primarily aim to facilitate communication between vehicles, we have Vehicular Ad-hoc Networks (VANETs) [2]. VANETs are networks that are self-organizing and comprised of vehicles. Research in the subject of communications is now seeing a surge in interest in vehicle communication. There are a lot of methods for vehicular communication these days, but IEEE 802.11p is where most people are putting their money. Among the various uses for VANET [3] are applications for life-critical and basic safety, group communication, internet access, electronic tall connection, and roadside service finding.

Figure 1.1 shows the Blackhole attack in VANET. Because VANET vehicles [4] are always on the go, routing in this network is no easy feat. It is possible for a rogue node to alter, delete, or reroute communications inside the network, or even completely divert traffic if it drops, blocks, or modifies messages. As a result, a safe framework for controlling the veracity and trustworthiness of communications must be developed. The whole system is vulnerable to certain types of routing attacks [5] . Furthermore, such assaults might reduce the network's performance. Since we've covered wormholes and grayholes before, let's move on to blackholes. In a blackhole attack, the malicious node will initially attempt to get other nodes to send packets via it by displaying the quickest path in its route reply. Next, it will patiently await the packet to arrive. Once it does, it will secretly drop the packet, creating the illusion of a black hole, while it is routed via the malicious node. With the use of Route Reply messages with fabricated optimum route data [6], the bad node in a blackhole attack lures other nodes into passing packets via itself. Reducing the number of hops shown may provide this type of optimality. Once the best route has been determined, other nodes in the network will be enticed to send data via the malicious node. An evil node may subtly provide the illusion of a black hole by dropping communications. In a blackhole, all it takes is one or more nodes to divert network traffic in the incorrect direction.

The need for vehicle communication has arisen as a result of recent developments in automotive technology. Vehicles that can communicate with each other and the roadside infrastructure must be equipped with intelligence. In major cities where traffic is a major issue, this technology will be lifesaver since it allows cars

*AI Advanced School, aSSIST University, 46, Ewhayeodae 2-gil, Seodaemun-gu, Seoul, Korea (sphong@assist.ac.kr))

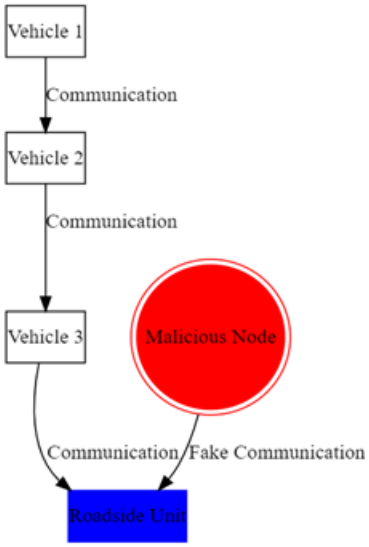

Fig. 1.1: Blackhole attack in VANET

to choose the best possible route based on the available information. By keeping themselves apprised of impending traffic conditions, drivers may choose the most efficient route, so conserving time, energy, and fuel. A variety of services depending on needs may be accessed by vehicles via connectivity with the infrastructure. Various researchers do a great deal of work for VANET [7] inside the context of an ad hoc network. The most common ”roadblock” to VANET technology is security. A VANET is meaningless without adequate security. The primary focus of VANET application governance is security management. VANET communication has to be protected from many kinds of attackers. Critical for VANET network security in the event of an attacker altering data contents, causing excessive latency, altering self-identity, or misbehaving in the network. Problems with centralised monitoring and security requirements, the open environment, and the high mobility of vehicles are limiting the adoption and expansion of VANET. The goal of this study is to identify several vulnerabilities in VANET adhoc networks and, using that information, to design and implement new safe methods that will provide greater protection against routing assaults, such as the Black Hole Attack [8].

The urgent need to guarantee the confidentiality and authenticity of data sent by Vehicular Ad Hoc Networks is the driving force behind this study (VANETs). VANETs are essential to contemporary transportation networks because they allow for the real-time interchange of data among vehicles, which improves safety and efficiency. Nevertheless, VANET settings are vulnerable to a variety of security risks due to their open and ever-changing nature, the most pressing of which being the potential of black hole attacks.

The realisation that safe communication is crucial to widespread use of VANETs in many contexts provides the impetus. The reliability of the VANET infrastructure is jeopardised by the possibility of black hole attacks, in which hostile nodes intentionally interrupt transmission by deleting or changing messages.

An effective hybrid method combining DBSCAN clustering with Decision Trees is the focus of the presented study, which intends to overcome this obstacle. The goal is to improve the efficacy and precision of detecting black hole attacks by combining the best features of the two methods. Decision Trees separate legitimate from fraudulent communication patterns among identified vehicle clusters using DBSCAN.

Our main objective is to find a way to make VANETs secure without lowering their performance or dependability. The need to improve the reliability of VANET connection by reducing the effects of black hole routing attacks is the driving force.

Results from experiments show that the suggested hybrid strategy is beneficial in improving packet drop rate, throughput, average end-to-end latency, and network routing load. In order to help create more secure and robust vehicle communication systems, this study is driven by the need to strengthen VANETs against new

threats.

The organization of paper is as follows; section 2 includes literature survey of Existing work; Section 3 includes methodology of proposed work; Section 4 includes experimental analysis of proposed work; section 5 includes conclusion and future work.

2. Literature Survey. Since 1970, research on adhoc networks has been underway. The original name for these networks was packet radio. Essentially, it's a way of thinking about setting up a short-term wireless network connecting nodes that are in motion. Because of how easy they are to use, MANETs and VANETs (Vehicular Adhoc Networks) are becoming more popular [9]. Compared to MANET, which tracks nodes via road infrastructure, VANET is superior. There are two main types of VANET communication: vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I). V2V communication refers to the exchange of data between vehicles equipped with On Board Unit (OBU) devices. When an OBU and an RSU exchange data, it's known as a V2I communication (Road Side Unit).

Various electromagnetic wavelengths, including infrared, microwaves, and radio waves, are used to carry out this communication. The VANET standard, developed by IEEE, is used in its implementation. The Wireless Access in Vehicular Environment (WAVE) standard, which is based on DSRC, is IEEE P1609.1 (Dedicated Short Range Communication). WAVE makes use of an updated version of IEEE 802.11a called IEEE 802.11p [10]. The guidelines for DSRC services, which use the 75 MHz spectrum between 5.850 and 5.925 GHz for both public safety and commercial purposes, were developed in 2003. One of the most important functions of the network layer is routing, which determines the best way to send data packets. The duty of the routing process lies with the routing protocols. Reactive and proactive routing are the two primary types of routing. From their unique vantage points, researchers have examined the problems, requirements, and priorities related to VANET security. Recent studies [11] have covered many forms of network assault and security measures to protect against them. In [12], the author provided a comprehensive overview of wireless adhoc networks and highlighted their security characteristics, privacy requirements, and shortcomings.

The author of [13] outlined the privacy and security issues that must be resolved before VANETs can be used in reality, and they also explained the communication architecture of these networks . The author discusses the difficulties with VANET security and the many assaults on VANETs in (9), and they categorise these assaults according to the various levels of VANET security. The author of in [14] discusses VANET security, including a thorough threat analysis and the best design for securing the network. Various security proposals put forward by different researchers are shown in [15]. The author surveyed current trust models in VANETs and addressed their main concerns. In order to achieve successful trust management in VANETs, the author also proposes desirable features. The author suggested a method for detecting Sybil attacks on VANET in [16]. The author outlined the current security standards and spoke about several ways to increase the vehicle's intelligence for better security in [15]. The author of [17] delves into the hierarchical structure of VANET and the many challenges it faces. A GPS time spoofing attack on a VANET was covered in [18].

The author of [17] outlined a VANET routing system for use in urban areas. Geographical forwarding is an attempt to enhance the routing process in urban traffic architecture. This study presents an evaluation of two routing protocols—Proactive and Reactive—using the simulator NS2.30 for a variety of city scenarios, and it details the inner workings of each. Some of the metrics used for result analysis include average delay, average delivery ratio, average route length, and network overhead. The results from [18] show that applications that are sensitive to throughput are better served by a reactive strategy, whereas applications that are sensitive to delays are better served by a proactive one. By demonstrating the research of several routing protocols, the author of this article explored the many obstacles of building routing protocols for VANETs. Various routing protocols were compared in this article. They broke the protocols down into five groups and spoke about each one: ad hoc, position-based, cluster-based, broadcasting, and geo-casting routing methods. The paper [19] provides an overview of several routing protocols. Security in mobile ad hoc networks and the many forms of attack on such networks are the topics of this paper. In this article, the author outlined the three pillars of network security: availability, confidentiality, and integrity. Various forms of attacks on ad hoc networks are covered, including active, passive, and advanced attacks. In this paper, we will only go over the many forms of attacks and the damage they may do to a network. Clustering and key distribution, efficient conditional privacy preservation, reputation checking, plausibility testing, and distributed key management are some of

the security mechanisms discussed in [20]. Based on the comparison provided, clustering and key distribution provide greater benefits than other accessible solutions.

Although there has been significant progress in Vehicular Ad Hoc Networks (VANETs) that might improve transportation systems' communication and safety, there is a clear paucity of study on how to tackle security issues, especially in relation to black hole attacks, in the current literature. To address the ever-changing nature of VANET systems, existing research either focuses on isolated approaches or fails to take a holistic view. Nobody has looked at the need for a strong hybrid system that detects black hole attacks by combining clustering and decision-making techniques.

Black hole attacks, in which hostile nodes deliberately discard or change messages, pose a growing danger to VANET security and may cause communication interruptions. There is a significant void in the creation of a dependable and efficient detection mechanism since current methods are either inaccurate or don't take VANETs' dynamic and resource-constrained characteristics into account.

Creating an all-encompassing solution that gets beyond the shortcomings of existing approaches is the present challenge. To be more precise, the task at hand is to develop a combined hybrid strategy that effectively detects black hole assaults in VANETs by combining the advantages of DBSCAN clustering with Decision Trees. All things considered, the success of VANET communication depends on a solution that improves security while also taking performance and reliability into account. To address this gap, the proposed study would provide a novel and efficient method to protect VANETs against the growing danger of black hole assaults.

3. Proposed Methodology. Because of the variety of assaults that may be launched in VANET, the role of the attacker is crucial. Attackers aim to disrupt other authorised users in order to cause difficulties in the operating environment. An attacker may alter the contents of a sent communication or delay or delete it entirely. Attacks against VANET might take several forms. Here, we mostly talk about routing attacks. Attackers mostly target weaknesses at the network layer in routing attacks. An attacker may disrupt the routing process and even lose packets in a routing assault. In this article, we will mostly cover routing attacks, which fall into three primary types: blackhole, wormhole, and grayhole. The initial step in a blackhole attack is for the malicious node to submit a route reply with the shortest path in order to lure other nodes into passing packets via itself. Once a rogue node has retrieved a packet from a specific node, it may covertly discard it, producing the illusion of a black hole. Figure 3.1 shows the Block Diagram of Proposed Methodology.

MF (Message Frequency), SSV (Signal Strength Variability), CC (Clustering Coefficient), AIT (Average Inter-Message Time), ND (Node Density), H (Entropy), FDM (Frequency of Messages Dropped/Modified), AFD (Anomalous Changes in Forwarding Decisions), SD (Sudden Disruptions in Communication Patterns), AEG (Alterations in the Connectivity Graph), UFC (Unusual Patterns in Claiming False Connectivity), DMP (Disruption in Paths for Message Transmission), IL (Increased Latency Caused by Manipulated Forwarding), EPL (Elevated Packet Loss Rates due to Black Hole Attacks). For Vehicular Ad Hoc Networks (VANETs) to effectively detect black hole assaults, a multi-stage process is necessary. The following are the main steps for detecting black hole attacks in VANETs:

1. Data Collection
   - Gathering data from the VANET environment, which may include real-world traces, simulated scenarios, or a combination of both.
   - Capture information such as communication logs, GPS traces, network parameters, and security-related metrics.
2. Preprocessing
   - Clean and preprocess the collected data to handle noise, missing values, and inconsistencies.
   - Transform the data into a suitable format for analysis.
3. Clustering using DBSCAN
   - Apply Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to group vehicles based on their spatial proximity and communication patterns.
   - Identify spatial clusters of vehicles, as anomalies within these clusters may indicate the presence of a black hole attack.
4. Feature Extraction
   - Extract relevant features from the clustered data that characterize normal and potentially mali-

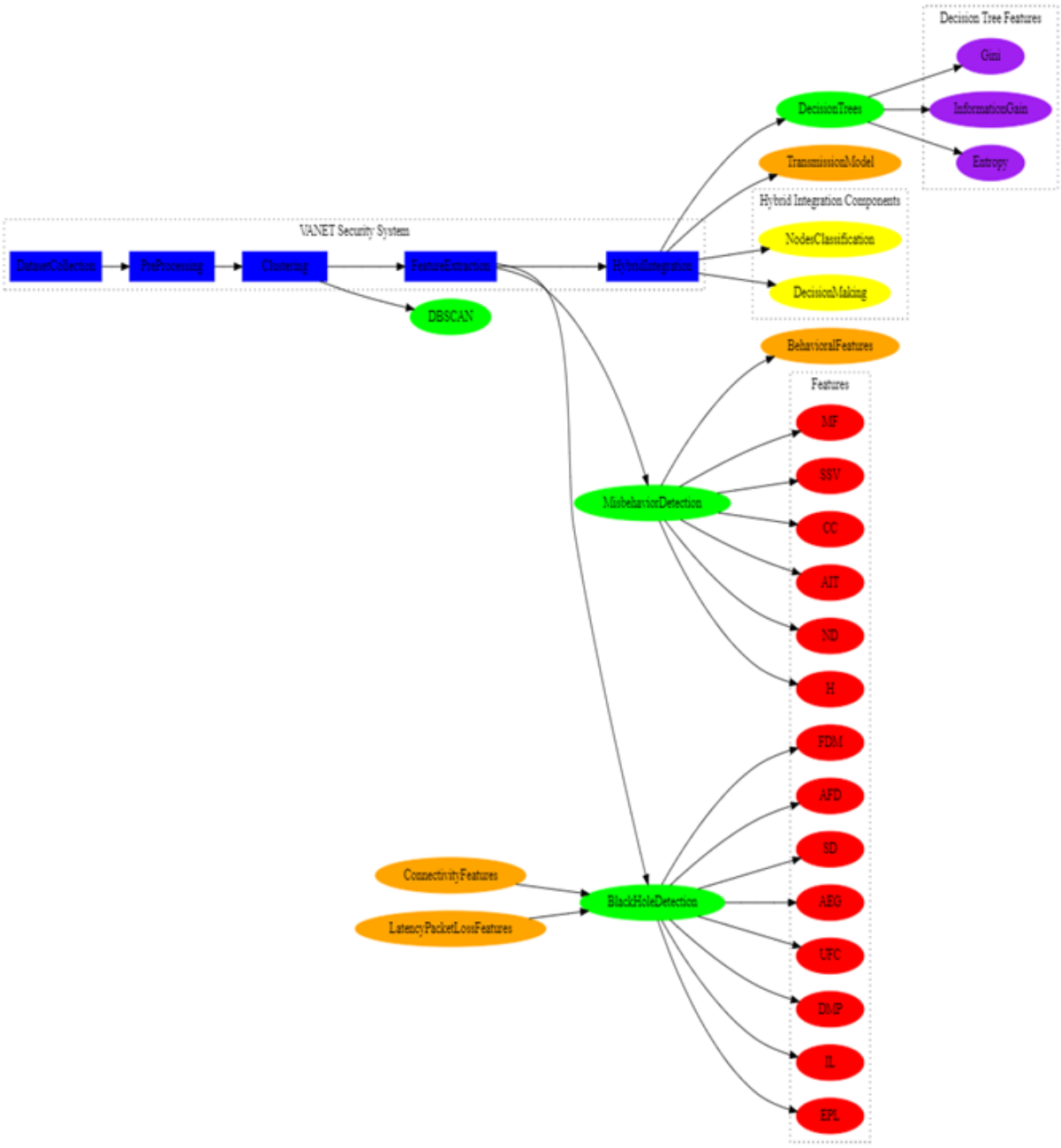

Fig. 3.1: Block Diagram of Proposed Methodology

cious communication patterns.

- Features may include metrics related to message frequency, signal strength, and node behavior within clusters.

5. Hybrid Integration:
   - Integrate the results of the clustering (DBSCAN) and classification (Decision Trees) stages to create a hybrid detection model.

- Develop a decision-making mechanism that considers the outputs of both components to enhance the overall accuracy and efficiency of black hole attack detection.

6. Performance Evaluation
   - Evaluate the performance of the hybrid approach using a set of predefined metrics, including Precision, Recall, F1 Score, Packet Drop Rate, Throughput, Average end-to-end delay, and Network Routing Load.
   - Compare the results against baseline models and individual techniques to assess the effectiveness of the proposed approach.

3.1. Dataset Collection. This study's data comes from an expansion of the VeReMi dataset, which is well-known in the field of Vehicular Ad Hoc Networks (VANETs). Using the Framework for Misbehavior Detection (F2MD), the dataset is carefully improved to include three key components. At its heart, the collection contains Cooperative Awareness Messages (CAM), which are crucial for depicting the data sent among VANET vehicles and include crucial elements like location, velocity, and direction. To further diversity the dataset and mimic harmful behaviours, a new class of assaults called the "Fake Reporting Attack" is established. This new kind of attack adds another degree of complexity to the dataset by causing rogue nodes to provide misleading information or fake reports. Another important component of this study is figuring out what the Fake Reporting Attack did and how it affected things, especially with regard to the virtual dangers that drivers confront. The purpose of this expanded and improved dataset, which was developed using a systematic and organised manner, is to provide a more thorough basis for investigating fraudulent activities, developing better detection methods, and strengthening the security resilience of VANETs.

3.2. Pre-Processing. Pre-processing is a crucial step in preparing raw data for analysis and Modeling. In the context of VANETs and misbehaviour detection, pre-processing involves several tasks such as handling missing data, normalization.

1. Handling Missing Data
   One common pre-processing task is addressing missing data, which can arise due to communication issues or other factors. Imputation methods, such as mean imputation or regression imputation, can be used to estimate missing values.

$$\hat{\mathbf{x}} = \Sigma_{j=1}^{n} x_j \tag{3.1}$$

   where $\hat{x}$ is the imputed value, x_j is the observed value, and n is the number of observed values.

2. Normalization
   Normalization ensures that features are on a similar scale, preventing certain features from dominating others. Min-max normalization is a common technique:

$$x_{norm} = \frac{x - min(X)}{max(X) - min(X)} \tag{3.2}$$

   where x_{norm} is the normalized value, x is the original value, min(X) is the minimum value in the dataset, and max(X) is the maximum value in the dataset.

3.3. Clustering using DBSCAN. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a robust clustering algorithm widely employed in various fields, including Vehicular Ad Hoc Networks (VANETs), due to its ability to discover clusters of arbitrary shapes and effectively identify outliers or noise points. The fundamental idea behind DBSCAN is to define clusters based on the density of data points within a specific neighborhood. Figure 3.2 shows the Flowchart of Proposed work.

The algorithm categorizes points as core points, border points, or noise points, depending on their connectivity and proximity to other points. A core point is one with a minimum number of neighbors within a specified radius, while a border point is within the radius of a core point but lacks sufficient neighbors to be a core point itself. DBSCAN proceeds to form clusters by linking density-reachable points and expanding the clusters until no more points can be added. This adaptability makes DBSCAN particularly suited for VANETs, where communication patterns may vary in density and exhibit non-uniform spatial distributions. The algorithm's ability to discern clusters based on the intrinsic density of the data contributes to its effectiveness in

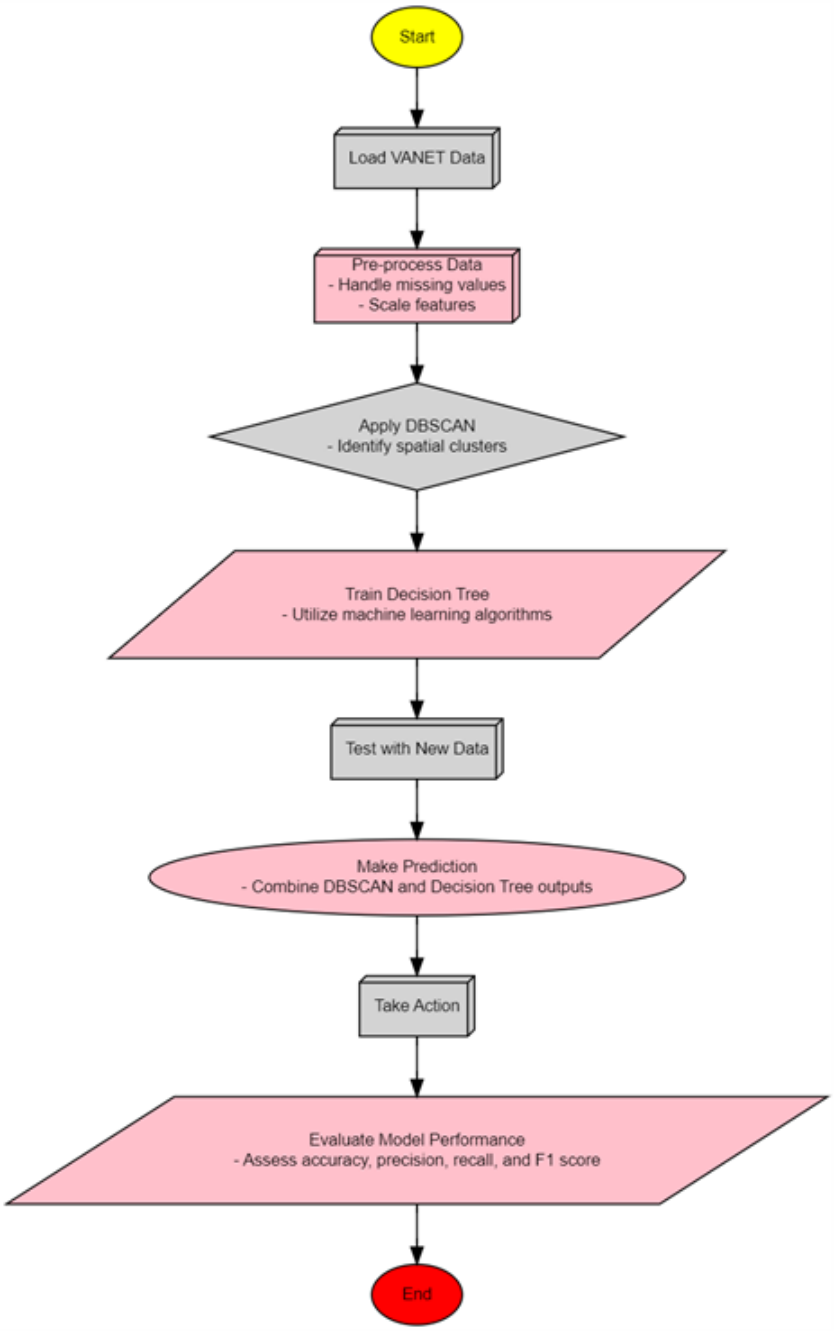

Fig. 3.2: Flowchart of Proposed work

uncovering meaningful structures in VANET communication, aiding in applications such as anomaly detection and misbehavior identification. DBSCAN categorizes data points into three types: core points, border points, and noise points.

Core Point (p): A point p is a core point if there are at least MinPts data points, including itself, within a distance of ϵ (a predefined radius). Border Point (q): A point q is a border point if it is within distance ϵ of a

core point but does not have enough neighbors to be a core point itself. Noise Point (s): A point s is a noise point if it is neither a core point nor a border point.

The reachability distance (r(p,q)) between two points p and q is the maximum of the core distance of p and the Euclidean distance between p and q.

$$r(p,q) = max(core_{distance}(p), ||p,q||) \tag{3.3}$$

The core distance ($core_{distance}(p)$) is the distance between a core point p and its MinPts-th nearest neighbor.

$$core_{distance}(p) = k_{distance}(p, MinPts) \tag{3.4}$$

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) employs a unique approach to form clusters by defining relationships between data points based on their density and proximity. Two critical concepts within DBSCAN are "Directly Density-Reachable" and "Density-Connected."

A point p is considered directly density-reachable from another point q if p falls within the reachability distance of q and q is a core point. This relationship is determined by comparing the core distance of q with the Euclidean distance between p and q. On the other hand, points p and q are density-connected if there exists a core point o such that both p and q are density-reachable from o. These definitions form the foundation for the DBSCAN algorithm.

The k-distance of a point p is the distance to its k-th nearest neighbor:

$$k_{distance}(p,k) = distance(p, N_k(p)) \tag{3.5}$$

$N_k(p)$ denotes the set of k-nearest neighbors of p, and distance (p,$N_k(p)$) is the Euclidean distance between p and its k^{th} nearest neighbor.

The DBSCAN algorithm starts by selecting an arbitrary point in the dataset and expanding the cluster by adding all directly density-reachable points to it. This process continues iteratively, encompassing additional points into the cluster until no more points can be added. The algorithm dynamically adapts to the varying density of the dataset, classifying each point as a core point, a border point, or a noise point. Core points initiate the expansion of clusters, while border points lie within the vicinity of core points but do not possess sufficient neighbors to be core points themselves. Noise points, lacking the density requirements, remain unassigned. The DBSCAN algorithm's effectiveness lies in its ability to uncover clusters of arbitrary shapes and efficiently identify outliers, making it well-suited for applications in VANETs where communication patterns exhibit diverse densities and spatial distributions.

3.4. Feature Extraction. In VANETs, feature extraction involves capturing distinctive characteristics from communication patterns, network parameters, and other relevant metrics. These features serve as input variables for machine learning algorithms or statistical models, aiding in the discrimination between normal and malicious behavior.

Message Frequency (MF). Represents the rate of message exchange within a specific timeframe.

$$MF = \frac{Number\ of\ Messages}{Time\ Period} \tag{3.6}$$

Signal Strength Variability (SSV). Captures the variability in signal strength, which may indicate the presence of malicious nodes interfering with communication.

$$SSV = Standard\ Deviation\ of\ Signal\ Strength \tag{3.7}$$

Clustering Coefficient (CC). Reflects the degree of connectivity within a spatial cluster of vehicles, identifying potential areas of interest.

$$CC = \frac{(2 \times Number\ of\ Actual\ Connections)}{Number\ of\ Possible\ Connections} \tag{3.8}$$

Average Inter-Message Time (AIT). Measures the average time between consecutive messages, helping to identify abnormalities in communication patterns.

$$AIT = \frac{Total\ Time}{Number\ of\ Messages - 1} \tag{3.9}$$

Node Density (ND). Quantifies the concentration of nodes within a specified region, providing insights into the spatial distribution of vehicles.

$$ND = \frac{Number\ of\ Nodes}{Area\ of\ Region} \tag{3.10}$$

Entropy (H). Measures the randomness or unpredictability of message distribution, assisting in detecting irregularities.

$$H = -\Sigma_{j=1}^{n} P(i) log_2 P(i) \tag{3.11}$$

These extracted features collectively create a descriptive and discriminative representation of the VANET communication environment. The inclusion of such features in the analysis enhances the accuracy of misbehavior detection models and contributes to a more comprehensive understanding of the VANETsystem dynamics.

3.5. Hybrid Integration. In VANETs, vehicles communicate with each other through wireless communication to share important information such as location, speed, and road conditions. The basic concept involves the transmission of Cooperative Awareness Messages (CAM) or other safety-related messages among neighboring vehicles. The propagation of a message can be represented mathematically, taking into account factors like transmission time and distance.

$$TransmissionDistance(d_{transmit}) : d_{transmit} = v.t_{transmit} \tag{3.12}$$

where v is the vehicle's speed, and $t_{transmit}$ is the transmission time.

Received Signal Strength(RSS).

$$RSS = \frac{P_t.G_t.G_r.(\lambda)^2}{(4\pi)^2.d^2} \tag{3.13}$$

where P_t is the transmitted power, G_t and G_r are the gains of the transmitting and receiving antennas, λ is the wavelength, and d is the distance between the antennas.

Figure 3.3 shows the process of message transfer and attack detection. Message Transfer and Attack Detection. In the proposed approach, black hole attack detection in Vehicular Ad Hoc Networks (VANETs) integrates the power of DBSCAN (Density-Based Spatial Clustering of Applications with Noise) for spatial clustering and Decision Trees for classification. Unlike the traditional method employing an SVM classifier, our approach enhances security by leveraging DBSCAN to identify spatial clusters of vehicles and Decision Trees to discern normal communication patterns from potentially malicious ones.

Firstly, DBSCAN is applied to group vehicles based on their spatial proximity and communication behavior. Nodes within clusters are categorized as core points, border points, or noise points. Border and noise points may indicate anomalies in the network, potentially signalling the presence of black hole attacks. Prediction of Black hole attacks in VANET depends upon behavioural, connectivity and latency, packet loss features.

3.5.1. Behavioral Features.

Frequency of Messages Dropped or Modified $F_{drop/modify}$. Count the occurrences of messages that are dropped or modified over a given time period.

$$F_{drop/modify} = \frac{Number\ of\ Dropped/Modified\ Messages}{Total\ Messages} \tag{3.14}$$

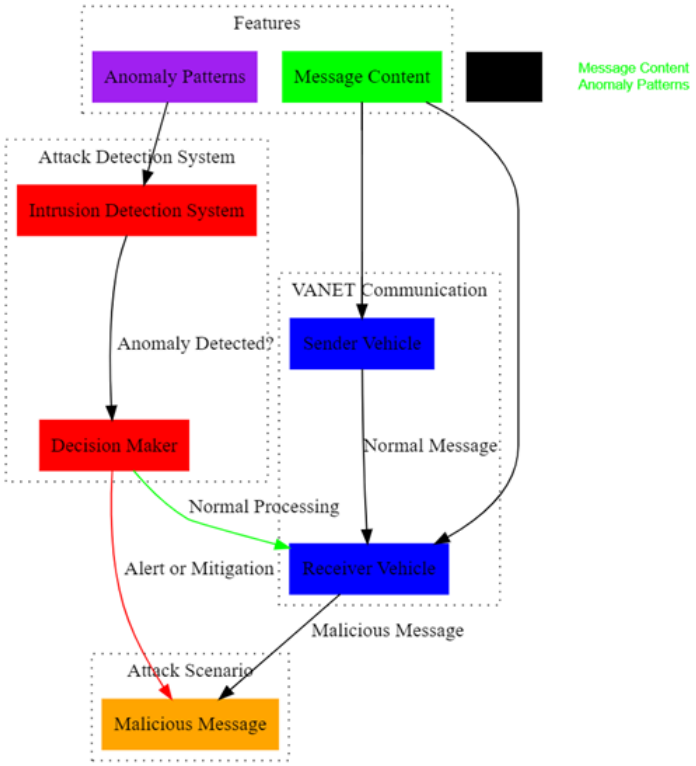

Fig. 3.3: Message Transfer and Attack Detection

Anomalous Changes in Forwarding Decisions Aforwarding. Measure unexpected alterations in the forwarding decisions of a node.

$$A_{forwarding} = \frac{Number\ of\ Anomalous\ Forwarding\ Decisions}{Total\ Forwarding\ Decisions} \tag{3.15}$$

Sudden Disruptions in Communication Patterns ($D_{disruption}$). Quantify abrupt changes in communication patterns, such as sudden stops in message transmission.

$$D_{disruption} = \frac{Number\ of\ Sudden\ Disruptions}{Total\ Communication\ Time} \tag{3.16}$$

3.5.2. Connectivity Features.

Alterations in the Connectivity Graph (A_{graph}). Evaluate changes in the connectivity graph by comparing the original and manipulated adjacency matrices.

$$A_{graph} = \frac{Number\ of\ Altered\ Edges}{Total\ Edges\ in\ Original\ Graph} \tag{3.17}$$

Unusual Patterns in Claiming False Connectivity ($U_{claiming}$). Identify abnormal claiming of false connectivity, indicating potential black hole attackers.

$$U_{claiming} = \frac{Number\ of\ Unusual\ Connectivity\ Claims}{Total\ Connectivity\ Claims} \tag{3.18}$$

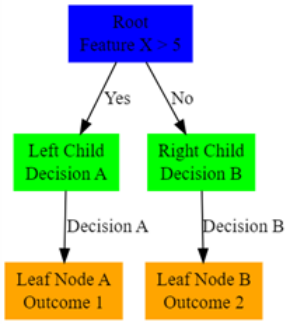

Fig. 3.4: Working of Decision Tree

Disruption in Paths for Message Transmission(D_{paths}). Assess the disruption in message paths due to false connectivity claims.

$$D_{paths} = \frac{Number\ of\ Disrupted\ Message\ Paths}{Total\ Message\ Paths} \tag{3.19}$$

Latency and Packet Loss Features:

Increased Latency Caused by Manipulated Forwarding ($L_{manipulated}$). Measure the average latency of messages when forwarding is manipulated.

$$L_{manipulated} = \frac{Latency\ of\ Manipulated\ Messages}{Number\ of\ Manipulated\ Messages} \tag{3.20}$$

Elevated Packet Loss Rates due to Black ↓ Attacks (P_{loss}). Calculate the packet loss rate when black hole attacks are suspected.

These equations provide a quantitative representation of the specified features, allowing for the assessment of abnormal behavior indicative of black hole attacks in VANETs. Figure 3.4 shows the Working of Decision Tree.

The thresholds for considering behavior as anomalous would depend on the specific characteristics of the VANET environment and the chosen detection strategy. Subsequently, the Decision Trees classifier is employed to classify nodes within the identified clusters. This step aims to differentiate between normal and potentially malicious communication patterns based on features extracted from the clusters. Decision Trees offer interpretability and the ability to capture complex decision boundaries.

$$Gini(t) = 1 - \Sigma_{i=1}^{c} p(i/t)^2 \tag{3.21}$$

where Gini(t) is the Gini impurity for node t; c is the number of classes; p(i/t) is the probability of class i at node t.

Information Gain measures the reduction in entropy or impurity achieved by splitting a dataset. For a split on feature A, the Information Gain (IG(A)) is calculated as follows:

$$IG(A) = H(parent) - \Sigma_j \frac{N_j}{N} H(child_j) \tag{3.22}$$

where H is the entropy, N is the total number of instances at the parent node, N_j is the number of instances in child node j, and H(parent) and $H(child)_j$ are the entropies of the parent and child nodes, respectively.

Entropy (H) is another measure of impurity. For a set of instances S, entropy is calculated as follows:

$$H(S) = -\Sigma_{i=1}^{c} p(i) log_2(p(i)) \tag{3.23}$$

Table 4.1: Simulation parameters for MDSR

Property	Value
Coverage Area	1000 m. X 1000 m
Number of Nodes	60
Simulation Time	600S
Transmission Range	250 m
Mobility	Random Way Point Model
Load	Data Payload 512 bytes.
Mobility Speed	20m/s
No of Gray hole Nodes	5
Connections	20 Pairs (40 nodes)
Traffic Type	UDP – CBR
Pause Time	0, 5, 10 and 15s
IDS Nodes	9 nodes (fixed)

where c is the number of classes, and p(i) is the proportion of instances of class i in set S.

The decision tree structure is built by recursively selecting features and thresholds to split the data. The decision-making process at each node involves choosing the split that maximizes information gain or minimizes impurity. The integration of DBSCAN and Decision Trees involves combining the results of these two stages. For instance, nodes classified as malicious by Decision Trees within clusters identified as anomalies by DBSCAN may be considered potential black hole attackers. The decision-making mechanism combines the spatial relationships identified by DBSCAN with the classification capabilities of Decision Trees to make a comprehensive determination of potential threats.

This hybrid approach enhances the accuracy and efficiency of black hole attack detection, providing adaptability to the dynamic and resource-constrained VANET environment. It gives a robust solution for discerning normal and malicious behavior, thereby ensuring the integrity of communication within the network.

4. Experimental Results and Analysis. This study used to verify that the suggested technique could effectively locate and isolate grey hole nodes. Within a 1000 m X 1000 m region, there are 50 normally behaving nodes that are using the MDSR routing protocol. There are also a few of bad nodes that are randomly placed and are selectively launching grey hole attacks. Additionally, there are a number of fixed IDS nodes. Each of the twenty pairings that were selected at random will be transmitting data at a rate of 5 kbps using UDP-Constant Bit Rate (UDP-CBR). A Random-way point model was used to move all the normal nodes at random rates ranging from 0 to 20 m/s. Furthermore, four distinct kinds of typical node stop times—0, 5, 10, and 15 seconds—were taken into account independently. The amount of time a mobile node may stay still before continuing to move is called its pause time. For instance, if the pause time is 0, it indicates that all nodes were moving continuously, without any brief pauses. The frequency of changes to the topology of a network is also indicated by the pause time. In Table 4.1, you can see the key parameters used in all of the Glomosim studies. The experimental data shown here is an average value derived from these 10 trials. Additionally, we compare our method to an existing one that was suggested follows a similar pattern to our method, with neighbour nodes of the source route doing monitoring and the source node sending data in blocks. It also doesn't use cryptography to identify threats.

We compare the proposed DBSCAN-DT framework's results to those of two other approaches already in use. Two algorithms that have been developed for use in WSN are the Adaptive Sink Aware (ASA) method and the Secure Route Discovery in AODV (SRD-AODV). Next, we use the table and graph values, in addition to the following metrics, to determine the performance of the proposed DBSCAN-DT framework in WSN.

4.1. Impact of Delay. In a WSN, the delay is defined as the time it takes for a data packet to travel from its source node to its destination node, and vice versa. Reducing WSN latency makes the suggested approach more efficient during transmission. Delay as a function of data packet count is seen in Table 4.2. The following table compares the suggested DBSCAN-DT framework to several current approaches, including SRD-AODV

Table 4.2: Tabulation for Delay

No.of data packets	Delay(ms)		
	Existing SRD- AODV	Existing ASA	Proposed DBSCAN-DT
10	53	47	41
20	55	49	43
30	56	50	44
40	62	56	50
50	64	58	52
60	70	63	58
70	73	66	61
80	71	64	59
90	76	69	64
100	77	70	65

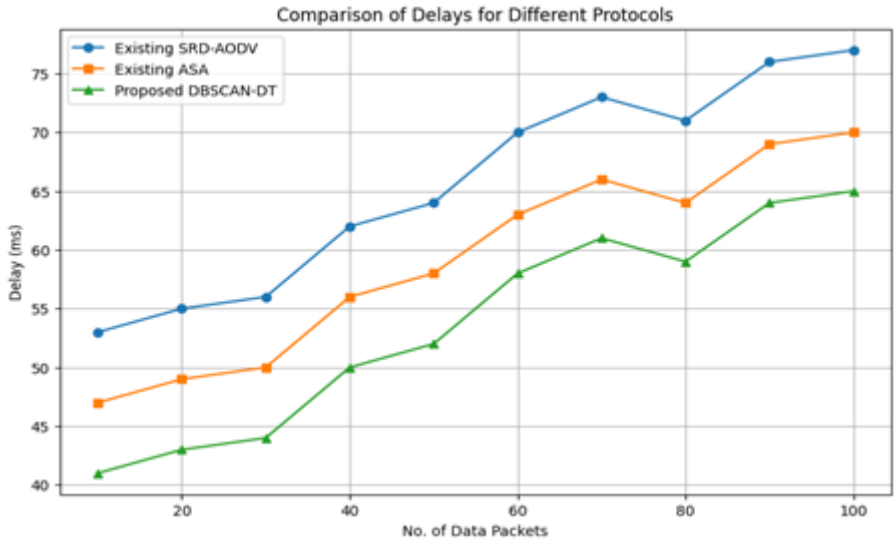

Fig. 4.1: Measure of Delay

[15] and the ASA algorithm [16]. For the purpose of conducting experiments, the quantity of data packets is adjusted between ten and one hundred. According to Table 4.2, all approaches experience an increase in latency while attempting to improve the amount of data packets.While the current approach for identifying blackholes and sinks in WSNs during secured transmission takes too much time, the suggested DBSCAN-DT architecture cuts down on that time significantly. Figure 4.1 shows the Measure of Delay.

The delay measurement for both the proposed and current approaches in WSN with different types of data packets is shown in Figure 6. Figure compares the suggested DBSCAN-DT framework to two current methods: SRD-AODV and the ASA algorithm. The numbers of data packets used for experimental analysis range from tens to hundreds. Consequently, the suggested DBSCAN-DT architecture, as opposed to current techniques, minimises latency in the sensor network. By using the suggested DBSCAN-DT structure, all patients' identical data are transmitted to the cluster head node, allowing for efficient evaluation of delay within cluster distances. By using the DBSCAN clustering technique, these nodes are able to accomplish the intrusion-measure correlation in WSN.

As a result, the correlation value is achieved by using just the most recent patient record based on its time, rather than passing the information of each node to the intrusion detection system. So, in comparison to the

Table 4.3: Tabulation for Attack Detection Accuracy

No. of sensor Nodes	Accurately identified attack node			Attack Detection Accuracy(%)		
	SRD-AODV	ASA	Proposed DBSCAN-DT	SRD-AODV	ASA	Proposed DBSCAN-DT
50	33	36	39	66	72	78
100	67	73	79	67	73	79
150	102	111	120	68	74	80
200	144	154	166	72	77	83
250	185	195	210	74	78	84
300	225	243	261	75	81	87
350	273	287	308	78	82	88
400	316	340	364	79	85	91
450	369	387	414	82	86	92
500	420	440	465	84	88	93

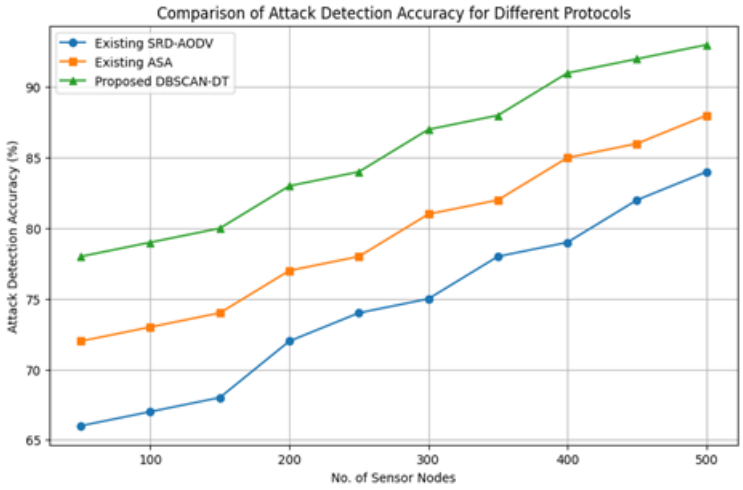

Fig. 4.2: Measure of Attack Detection Accuracy

current state-of-the-art algorithms, the suggested DBSCAN-DT framework significantly reduces data packet delivery delays in WSNs by 19% compared to the SRD-AODV and by 10% compared to the ASA algorithm.

4.2. Impact of Attack Detection Accuracy. The attack detection accuracy is the percentage of sensor nodes in a WSN that correctly identify attack nodes, sometimes called sinkholes or black holes. The following is a mathematical representation of the attack detection accuracy. The suggested method promises to be more effective if the network's attack detection accuracy is enhanced.

Based on the number of sensor nodes, Table 4.3 shows the attack detection accuracy utilising the proposed DBSCAN-DT framework and current approaches, such as SRD-AODV and the ASA algorithm . The experimental work is carried out using a range of 50 to 500 sensor nodes. Increasing the number of sensor nodes improves the accuracy of attack detection for all approaches, according to the values in the table.

Accuracy in detecting attacks as a function of sensor node count is shown in Figure 4.2. In addition, the figure compares the proposed DBSCAN-DT framework to two existing algorithms, the ASA method. As a result, the suggested DBSCAN-DT frameworkas improves the accuracy of attack detection when compared to the current methodologies. The suggested DBSCAN-DT framework effectively improves the accuracy of attack

Table 4.4: Tabulation for Packet Delivery Ratio

No.of data packets	No.of data packet received			Packet delivery ratio(%)		
	SRD-AODV	ASA	Proposed MK-Means	SRD-AODV	ASA	Proposed MK-Means
10	7	7	8	71	74	83
20	14	15	17	73	75	85
30	22	24	26	75	80	87
40	30	33	35	78	83	88
50	39	42	44	79	84	90
60	47	51	54	81	85	91
70	58	61	64	83	87	93
80	66	72	74	85	90	94
90	77	82	85	87	91	96
100	87	92	96	71	83	92

detection for secured broadcasting using a machine learning approach. The document outlines the intrusion measure that is used to confirm attacks. Additionally, the value of the intrusion measure is a growth value that is tied to time, which helps in retrieving patient information about a given moment. Not only that, the suggested DBSCAN-DT framework offers improved accuracy in identifying attacks quickly by using time-related growth value to determine whether an attack is noticed for a normal node. Consequently, the proposed DBSCAN-DT framework outperforms the state-of-the-art SRD-AODV by 15% in WSN and the state-of-the-art ASA method by 7%.

In contrast to the current methodologies, the suggested DBSCAN-DT architecture reliably identifies SH and BH attack nodes to provide safe delivery via a WSN. The data in the table below Figure 7 is used to produce the graph.

4.3. Impact of Packet Delivery Ratio. The packet delivery ratio, as calculated using the suggested DBSCAN-DT framework, is the percentage of data packets that reach their intended recipients without error out of all the data packets sent over the network. What follows is a mathematical depiction of packet delivery ratio.

Based on varying data packet counts, Table 4.4 shows the experimental results of the packet delivery ratio for the current technique and the suggested one. Table 4.4 shows that all techniques have an enhanced packet delivery ratio as the number of data packets increases.

Figure 4.3 shows the packet delivery ratio measured using Ghazaleh Jahandoust and Fatemeh Ghassemi's ASA algorithm and current approaches. The amounts of data packets used for experimental analysis range from tens to hundreds.

Figure 4.4 shows that the suggested DBSCAN-DT framework outperforms the current approaches in terms of packet delivery ratio in WSN.

The suggested DBSCAN-DT method is the basis for this significant enhancement in the packet delivery ratio. The next step is to take into account rescaled entity points that include various patient characteristics (this is because patient data are not static). Nevertheless, when contrasted with other current approaches, the suggested DBSCAN-DT framework considerably improves the packet delivery ratio during data packet transmission from source node to sink node in the network. Figure 4.3 provides the data used to generate the graph. The MINRMAXR method effectively reduces packet loss in WSN attack detection by enabling rescaled entity points and non-overlapping subsets. As a result, the rate of attack detection using the retrieved characteristics is improved. Thus, the suggested DBSCAN-DT structure helps achieve a greater packet delivery ratio.Table 4.5 shows the Tabulation for Computational Complexity

Nevertheless, when contrasted with other current approaches, the suggested DBSCAN-DT framework considerably improves the packet delivery ratio during data packet transmission from source node to sink node in

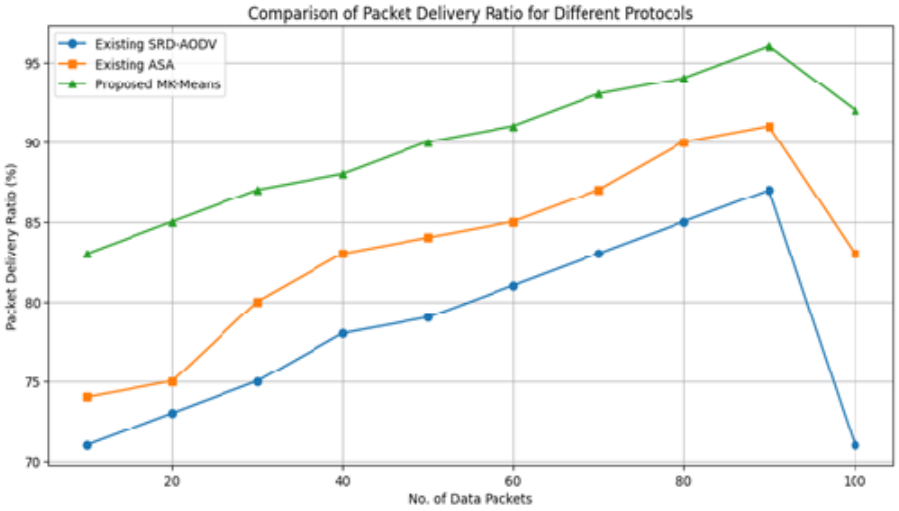

Fig. 4.3: Measure of Packet Delivery Ratio

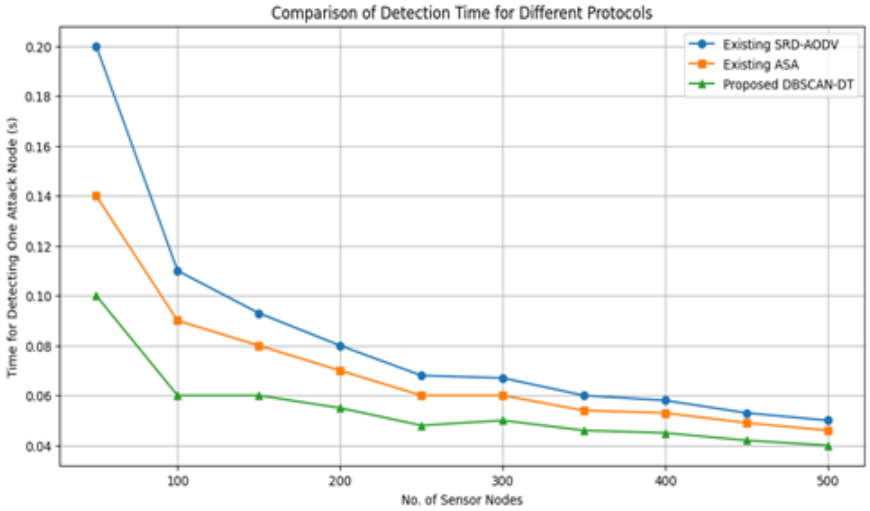

Fig. 4.4: Measure of Detection Time

the network. Figure 4.4 provides the data shows the detection time and Figure 4.5 shows the computational complexity of the proposed work.

5. Conclusion. In order to reduce computational complexity in WSN and improve attack detection accuracy, the DBSCAN-DT framework is developed. Three procedures, including physiological data collection (PDC), proportional overlapping score (POS), and machine learning approach, make up the proposed DBSCAN-DT system. The PDC module starts by collecting features from the training dataset, which are based on physiological parameter measurements. The next step is to use the POS model for data pre-processing and feature minimization on the chosen training data. This simplifies the task of detecting attacks while they are being sent. The next step is to use a wireless communication network to send the chosen characteristics to the DBSCAN clustering algorithm, which will then conduct the testing and classification. The detection metrics include the number of packets transmitted and received, which aid in the calculation of Intrusion Measure

Table 4.5: Tabulation for Computational Complexity

No. of sensor nodes	Time for detecting one attack node			Computational Complexity (ms)		
	SRD-AODV	ASA	Proposed DBSCAN-DT	SRD-AODV	ASA	Proposed DBSCAN-DT
50	0.2	0.14	0.1	10	7	5
100	0.11	0.09	0.06	11	9	6
150	0.093	0.08	0.06	14	12	9
200	0.08	0.07	0.055	16	14	11
250	0.068	0.06	0.048	17	15	12
300	0.067	0.06	0.05	20	18	15
350	0.06	0.054	0.046	21	19	16
400	0.058	0.053	0.045	23	21	18
450	0.053	0.049	0.042	24	22	19
500	0.05	0.046	0.04	25	23	20

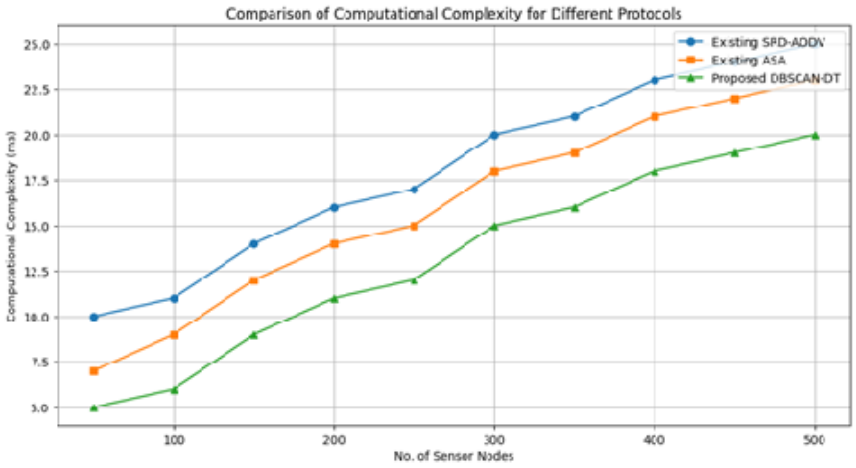

Fig. 4.5: Measure of Computational Complexity

(IM) utilising IDS, using the suggested DBSCAN-DT architecture. As a result, it can quickly and accurately identify the sinkhole or black hole attack node, allowing for quick and secure communication. By alerting the network to stop data transmission in the event that a SH or BH assault is detected in a WSN, intrusion detection systems improve the accuracy of attack detection. Improved packet delivery is therefore a benefit of the suggested DBSCAN-DT system. proportion with enhanced efficacy. Additionally, measures such as computational complexity, latency, packet delivery ratio, and attack detection accuracy are used to evaluate the performance of the proposed DBSCAN-DT framework. In comparison to state-of-the-art studies, the simulation results show that the suggested DBSCAN-DT framework considerably improves performance by reducing computing complexity, improving packet delivery ratio, and increasing attack detection accuracy. Further, it reduces latency.

6. Acknowledgment. This research was supported by AI Advanced School, aSSIST University, Seoul, Korea.

REFERENCES

[1] Akashdeep Sharma et al, *IoT and deep learning-inspired multi-model framework for monitoring Active Fire Locations in Agricultural Activities*, Computers and Electrical Engineering, vol. 93, 2021. ,
[2] Abdelmaguide et al, *VeReMiAP: A VeReMi-based Dataset for Predicting the Effect of Attacks in VANETs*, 2023.
[3] Hasrouny et al, *VANet security challenges and solutions: A survey*, Vehicular Communications, vol.7, 7-20, 2017.
[4] Lee M et al *Vanet applications: Past, present, and future*, Vehicular Communications, vol.28, 100310,2021.
[5] Cooper et al *A comparative survey of VANET clustering techniques*, IEEE Communications Surveys and Tutorials, vol. 19(1), 657-681, 2016.
[6] Hasrouny et al *VANet security challenges and solutions: A survey*, Vehicular Communications, vol. 7, 7-20, 2017.
[7] Mishra R et al *VANET security: Issues, challenges and solutions*, In 2016 international conference on electrical, electronics, and optimization techniques (ICEEOT), IEEE, 1050-1055, 2016.
[8] Mejri M.N et al *Survey on VANET security challenges and possible cryptographic solutions*, Vehicular Communications, vol. 1(2), 53-66, 2014.
[9] Mansour M.B et al *VANET security and privacy-an overview*, International Journal of Network Security and Its Applications (IJNSA), Vol. 10, 2018.
[10] Karabulut M.A et al *Inspecting VANET with various critical aspects–a systematic review*, Ad Hoc Networks, 103281, 2023.
[11] Al-Shareeda M.A et al *A Systematic Literature Review on Security of Vehicular Ad-hoc Network (VANET) based on VEINS Framework*, IEEE Access, 2023.
[12] Gaba S et al *A comprehensive survey on VANET security attacks*, In AIP Conference Proceedings, Vol. 2495(1), 2023.
[13] Abdulkadhim F.G et al *Design and development of a hybrid (SDN+ SOM) approach for enhancing security in VANET*, Applied Nanoscience, vol. 13(1), 799-810, 2023.
[14] Piramuthu O.B et al *VANET authentication protocols: security analysis and a proposal*, The Journal of Supercomputing, Vol. 79(2), 2153-2179, 2023.
[15] Mahi M.J.N. et al *A Review on VANET Security: Future Challenges and Open Issues*, Indonesian Journal of Electrical Engineering and Informatics (IJEEI), Vol.11(1), 180-193, 2023.
[16] Narayanan K.L et al *An efficient key validation mechanism with VANET in real-time cloud monitoring metrics to enhance cloud storage and security*, Sustainable Energy Technologies and Assessments, Vol.56, 102970, 2023.
[17] Mdee A.P et al *Security compliant and cooperative pseudonyms swapping for location privacy preservation in VANETs*, IEEE Transactions on Vehicular Technology, 2023.
[18] Gnanajeyaraman R et al *VANET security enhancement in cloud navigation with Internet of Things-based trust model in deep learning architecture*, Soft Computing, 1-12, 2023.
[19] Raut R.M et al *A Survey on Security Threats in VANET and Its Solutions*, In International Conference on Recent Trends in Artificial Intelligence and IoT, pp. 229-240, 2023.
[20] Rajeswari R.M et al *Enhance Security and Privacy in VANET Based Sensor Monitoring and Emergency Services*, Cybernetics and Systems, 1-22, 2023.
[21] Velayudhan et al, *Sybil attack detection and secure data transmission in VANET using CMEHA-DNN and MD5-ECC*, Journal of Ambient Intelligence and Humanized Computing, Vol.14(2), 1297-1309, 2023.

Edited by: Mahesh T R
Special issue on: Scalable Dew Computing for future generation IoT systems
Received: Jan 30, 2024
Accepted: May 9, 2024