

Received 30 May 2023, accepted 23 June 2023, date of publication 12 July 2023, date of current version 26 July 2023. *Digital Object Identifier* 10.1109/ACCESS.2023.3294469

RESEARCH ARTICLE

LCSS Based Sybil Attack Detection and Avoidance in Clustered Vehicular Networks

S. RAKHI¹⁰¹, (Member, IEEE), AND K. R. SHOBHA¹⁰², (Senior Member, IEEE)

¹Atria Institute of Technology, Bengaluru, Karnataka 560024, India ²Ramaiah Institute of Technology, Bengaluru, Karnataka 560054, India Corresponding author: S. Rakhi (rakhi.s@atria.edu)

ABSTRACT Future Road transportation mainly depends upon connected vehicles. Intelligent Transportation Systems bring benefits to the road users through Vehicular Adhoc Networks (VANETs). Since VANET packet contains life critical information, security is inevitable. A rogue node called sybil node can transmit fake messages to its neighbours and disrupt the system, challenging security. Since the nodes are very dynamic, stability is also a major concern. Existing rogue node detection methods do not address this problem suitably. In the proposed work, rouge node detection is implemented in a clustered network which improves the stability of the network. The main aim of this paper is to implement a sybil attack detection method in distributed or coordinated clustered networks using a novel hybridization technique. The cluster head detects the sybil attacker by comparing the received signal strength of packets from each node based on a similarity algorithm, Longest Common SubSequence (LCSS). However, if the sybil attacker launches a power control mechanism, the similarity calculation fails. To overcome this issue, a Change Point Detection(CPD) technique by comparing the changes in mean value of RSS time series from a particular node is proposed. Coordinated attacks can be easily detected in a clustered network as the information regarding the attackers' spreads in the network quickly so that the nodes can avoid connecting to such malicious nodes during their journey. The proposed algorithm shows significant improvement in detection rate, detection delay and false positive for varying vehicle count compared to existing techniques.

INDEX TERMS Cluster, power control, RSSI, Sybil, VANET.

I. INTRODUCTION

Vehicular ad hoc network (VANET) is a key factor of Intelligent Transportation System. VANET is definitely going to make our lives safe on roads in the future. Connected vehicles have become reality after the deployment of Google cars on roads. The main feature of VANET is the selforganization of vehicles which helps in information sharing and quick communication. V2V (Vehicle to Vehicle) and V2I (Vehicle to Infrastructure) communication in VANET is based on Dedicated Short Range Communication (DSRC) standard. VANET has the ability to make the hazardous roads safer to the passengers. The communication between the vehicles is mainly based on the information exchanged between the vehicles on road. VANET is an adhoc wireless network, hence it inherits all threats in wireless medium

The associate editor coordinating the review of this manuscript and approving it for publication was Wentao Fan^(D).

resulting in a major concern for security. One of the major attacks in VANET is sybil attack. The sybil attack is one of the harmful attacks in VANETs because it severely damages the security of network which leads to a threat to the lives of drivers and passengers. Sybil attack is a security threat where the attacker creates multiple identities and spreads data in the network to create confusion. It is an active, insider, rational attack. It is easy to launch any other attacks like blackhole and denial of service attack in the network if a sybil attacker is launched in the VANET. Sybil attack can cause traffic jams, accidents and chaos in the network. Hence, it is always necessary to guard the network against malicious activities. Several security schemes are proposed to detect and avoid sybil attacks [1].

It is easy for an intruder to deploy an attack in the highly mobile environment. Therefore, detecting the presence of sybil attack and finding the sybil nodes become a challenge in transient neighbourhood. There are mainly four broad categories of sybil attack detection techniques namely resource testing based, cryptography based, trust based and physical measurement based [2]. Resource testing-based techniques fails if the sybil attacker has more computational resources. Cryptographic techniques are based on security key exchange between a central authority and vehicles. This technique is mainly based on centralized infrastructure and complex cryptographic key exchanges. In trust-based approach, node calculates trust value based on parameters like number of neighbours, link lifetime, number of packets received from those nodes etc. These methods may not be suited for a high mobility scenario. If we consider the decentralized nature and simpler technique, physical measurement approaches are well suited in VANETs. Received Signal Strength Indicator (RSSI) of beacon packets is the major component in detecting sybil attacks in this approach.

Stability is a major concern in dynamic scenario. Clustering vehicles has proven to improve the stability and reliability of the network in the past. A priority-based data centric clustering approach is implemented in the proposed work to increase stability of the network and the sybil attack detection algorithm is executed in the cluster heads. In a stabilized clustered network, it is easier to disseminate the messages in a wider range. Once an attack is detected, this information reaches all the nodes at a faster rate than in the unclustered network. The RSSI based sybil detection usually try to locate the nodes based on the distance for sybil detection [3]. In the proposed method comparison of the RSSI time series of nodes is considered. The fabricated nodes by the sybil attacker transmits similar RSSI time series since the RSSI calculation is mainly based on distance. Cluster heads compare the RSSI time series of all nodes and if the time series is similar in few nodes they are considered as sybil nodes. RSSI methods proposed earlier does not take care of deliberate power control performed by attackers. This work proposes a change point detection method to overcome the issue. When more than one attacker coordinates with each other it can have a complete control over the network. In a clustered environment, cluster head can easily detect the coordinated attacks since the detection is purely based on similarity calculation. The proposed sybil attack detection algorithm is compared with Multichannel Sybil attack detection algorithm called Voiceprint [4] which uses received signal strength indicator to detect attackers.

The contribution of the proposed work is summarized as follows:

- The proposed work is implemented with a novel hybridization technique using LCSS and CPD to detect sybil attackers in a coordinated or distributed network.
- A new sybil attack detection method is proposed using received signal strength calculation to identify rogue nodes in a clustered network.
- A modified LCSS algorithm is proposed to find the RSSI similarity index, to detect the sybil nodes in clustered network.

- To detect attackers performing power control, change point detection method is incorporated. The data collected are analysed using statistical techniques before taking decision.
- The cluster heads inform the neighbours regarding the attackers, thus the nodes can avoid accepting beacons from them during their journey.

The rest of the paper is organized as follows. Section II describes related work on sybil attack detection and clustering. Section III explains the proposed detection method in detail. Section IV has simulation results and analysis to evaluate proposed approach. Section V draws the conclusion.

II. RELATED WORK

Sybil attack was first described by Doucer [5]. In this attack an attacker creates multiple nodes in the network called sybil nodes. Sybil attack can cause a minor traffic jam to major accidents targeting human life. Hence, detection of sybil attack is very important as we are heading towards autonomous vehicles in the near future. To detect sybil attacks many methods are proposed recently. The physical measurement based position verification is a commonly used method based on finding the physical locations of nodes and comparing with claimed position. Radio resource testing is mainly dependent on the receiver and transmitter antennas and related radio modules. But if attackers are having multiple radio modules it is difficult to detect attackers. In public key cryptography techniques nodes exchange digital certificates and encryption methods with a central authority. But this method needs complete infrastructure support and overhead is more. In trust-based techniques the nodes reputation value is calculated and compared with neighbouring nodes in the network. Trust between the vehicles is important in this case.

Sybil detection methods using received signal strength indicator or power received based on absolute position or relative distance have been implemented by researchers in the recent past [6], [8], [16], [17], [18]. RSSI based method is basically light weight without any specialized hardware. But it mainly depends on the radio propagation model used in the system. Multichannel Sybil attack detection algorithm called Voiceprint [4] is based on RSSI time series as the vehicle's speech is used to compare the similarity among the time series. Time series is a sequence of data points consecutively collected over a time period. RSSI time series of two sybil attackers show very similar patterns during a time period. This technique does not depend on radio propagation model or Road Side Unit (RSU). It follows a data centric approach. It neither depends on the information from neighbouring vehicles nor support from RSU. Authors have experimented this algorithm in real scenarios. But this method fails if the attackers exhibit power control mechanism and also it is time consuming.

Baza et al. proposed a proof of work based algorithm for detection of sybil attacks. Each RSU issues a signed timestamped tag as a proof for the vehicle's location [7]. Upon receiving the proof of location from an RSU, the vehicle must solve a computational puzzle. It should provide a valid solution to the next RSU before it can obtain a proof of location. This prevents the vehicles from creating multiple IDs. Detection rate is more in this method. But continuous infrastructure support is required. Garip et al. proposed a RSSI based localization method called INTERLOC which estimates not only a location but an entire area even when GPS is unavailable [8]. This work continuously studies the change in interference level and adjusts itself to improve accuracy in localization. This method is robust to changes in interference levels and does not depend on RSU. It uses interference aware shadowing radio model. The method provides high accuracy. However, this method requires coordination from other nodes to localize the area.

Tulay et al. [9] proposes a method based on channel state information profiles. The Channel State Indicator (CSI) values for a node is estimated using the preambles at the beginning of each packet broadcasted from the node. CSI values are calculated from subcarriers. Pearson correlation coefficient between CSI of different nodes should be low if they are not at the same location. If the correlation coefficient is above a certain threshold, these nodes are Sybil attacker. Chen et al. [10] introduced a Sybil detection technique which takes the help of centralized infrastructure to monitor and record the RSSI values from each node. The centralized unit calculates the similarity of RSSI of node pairs and detects sybil attackers. This technique cannot be applied in a decentralized system. K Means algorithm is used to derive the test statistics for detection. Detection of power control of nodes are also addressed by the authors. Euclidian distance is used to find the similarity. So, this method can be used only if the number of sequences obtained are same from each node, which cannot be possible in a real time vehicular network.

Power control identification, PCISAD [3] is a Sybil attack detection technique when Sybil nodes deliberately alter the transmission power and RSSI which can lead to imprecise localization. In such cases it is difficult to differentiate Sybil nodes from normal nodes. In this method RSSI is calculated in both Common Control Channel (CCH) and Service Channel (SCH) channels. Sybil nodes are identified using Dynamic Time Warping (DTW) and power control is detected by Pruned Exact Linear Time (PELT) technique. DTW distance is the total accumulated cost of optimal warp path. PELT is to estimate the time at which the property of a RSSI time series changes. This method does not depend on any propagation model. The performance of PCISAD to detect Sybil attacks is much better than other similar methods like Voiceprint. The true positive and false rates give a better value. However, DTW based method requires high computational time and are sensitive to outliers. Murat et al [6] proposed a sybil attack detection using RSSI in WSN. A particular node uses the ratio of RSSIs of multiple nodes at different times to localize a particular node. If multiple ids are corresponding to same locations, they are Table 1. shows a comparison of various RSSI based detection techniques. Recently, the detection of VANET attacks has made substantial use of machine learning methods. Authors in two-layer collaborative IDS [11] propose a two-layer machine learning based IDS system to detect collaborative attacks in VANET. A collaboration between nodes called On board Vehicle IDS (OVIDS) and the mobile edge node server (MEC) is required in this method as two layers. OVIDS module works in tandem with MEC to provide effective detection. RSU acts as a gateway to edge servers that are very near to the nodes. This reduces the end-toend delay. Two KPIs considered here are latency and energy consumption. All data from the neighbours are aggregated by OVIDS, and normal and abnormal nodes are classified based on a bi-class classifiers in machine learning. Suspicious data are transmitted to Edge IDS(EIDS) in the MEC server for further analysis. It alerts the base station regarding the attack. Any coordination of the attacks can be avoided by the data aggregated by the OVIDS. The method has very high accuracy and fewer false positives with a reduced latency of detection. A unique majority voting-based collaborative architecture is suggested to identify the Sybil attack in the network [48]. The framework operates by simultaneously assembling various classifiers, such as K-Nearest Neighbour, Naive Bayes, Decision Tree, SVM, and Logistic Regression. For a final prediction, the Majority Voting (Hard and Soft) technique is used. 95% accuracy is claimed by the authors.

considered to be attackers. This is done by a cluster head. The

detection rate is high in this algorithm. If the attackers change

the transmission power deliberately, then this method fails.

Algorithm	Propagati on	Centrali zed(C)/	Infrastruct ure	Environme nt
	Model	Decentra	Support	
		lized(D)	Required	
Multi	Emperical	D	No	Dynamic
channel				
Interloc	Shadowing	С	Yes	Dynamic
K-means	Shadowing	D	No	Dynamic
clustering				
PCISAD	Emperical	D	No	Highly
				Dynamic
Wang	Jakes	D	No	Static
Murad	Freespace	D	No	Static

TABLE 1. Comparison of RSSI based detection algorithms.

Clustering is a method for grouping vehicles that have similar parameters. There are various clustering techniques in VANETs [12], [15]. In [13] Cluster head (CH) detects the attacker by calculating the difference in the power received at instant t and t+1. If the difference is above a threshold level, then the node is detected as a sybil. Cluster members performs localization technique to check if CH is an attacker. Even though authors claim to detect 92% attackers, this method assumes constant transmitted power, which is not possible in a real time scenario. Zang et al. [14] proposes a clustering using AODV based on edge computing strategy RSU is considered as edge node/CH. A reward function is calculated by CH with hop number HOP, link holding time T between vehicle, neighbour vehicle nodes, and energy consumption using Q learning process. Q learning gives optimal value. CMs are chosen based on highest reward.

Although few of the above methods overcome the adaptive power control techniques by the Sybil nodes, it is still a challenge in dynamic vehicular environment. Since the received RSSI values change with respect to transmitted power, detection of sybil using RSSI is a major challenge. Most of the Sybil detection methods assumes that nodes transmit messages with fixed power. The existing techniques for sybil attack detection use additional hardware or complex cryptographic solutions. A novel method is proposed here by combining clustering technique in detection of sybil attack.

In the proposed method a clustering technique is used to improve the stability of the network and to detect any kind of coordinated and uncoordinated attack. CH detect sybil nodes by calculating the similarity index of RSSIs between a pair of nodes in a cluster. Since CH detects the attacker in its cluster, the information to the other members and neighbouring CHs can be broadcasted faster. The number of nodes receiving information regarding the attackers are more in a less time compared to other methods.

III. SYSTEM DESIGN

In this section, the proposed system is explained in detail. The sybil detection method is performed by two different techniques named Longest Common Subsequence (LCSS) technique and Change Point Detection (CPD) method. The design is implemented in a clustered environment. LCSS algorithm is for RSSI similarity calculation and CPD is used for finding any transmitted power variations. The anomaly detection can be performed at a higher rate using LCSS algorithm in a clustered network [19], [20]. Sybil detection without power control can be easily detected by this approach. But if the node performs power variation during transmission this method cannot give accurate results. In order to overcome this issue a change point detection method is implemented. Both methods are performed by the CH to detect the sybil nodes. RSSI time series of Sybil nodes exhibit similar patterns since RSSI is based on distance. The RSSI can be easily calculated based on the following equations. Pt is the transmitted power, P(do) is the power calculated at a reference distance $d_0 = 1m$., α is the path loss constant and X_{σ} is the random variable.

$$RSSI = Pt - (P(d_o) + 10\alpha \log 10(d/d_o) + X_{\sigma})$$
(1)

$$P(d_o) = (Pt.Gt.Gr\lambda^2) / (4\pi^2 Ld_o^2)$$
⁽²⁾

To implement the architecture following assumptions are made.

• Network is clustered using the method, priority based clustering technique(MPMC) discussed in next section.

- Radio modules in each vehicle in the network follow DSRC standard.
- Malicious nodes are 5-20% of total nodes.
- Legitimate nodes do not perform power control.
- Sybil attackers perform power control during communication.



FIGURE 1. System architecture.

A. CLUSTERING ALGORITHM

Clustering algorithms helps in stability and load balancing in VANETs. The system architecture is shown in in Figure 1. Consider a road with 6 vehicles. Node 1 is the cluster head and 2-5 are direct members and node 6 is a multihop member. The sybil detection algorithm is implemented at each cluster head. In this work a multihop clustering algorithm called Modified Priority Based Multihop Clustering (MPMC) is explained which is a part of this implementation and an extension of the work (PMC) algorithm [12]. The MPMC algorithm has shown better stability compared to the PMC algorithm. Once the network is deployed, nodes move independently and choose the best parent to follow thus initiating the clustering process in the network. Every vehicle maintains a state in the cluster during this period. Cluster head selection is an important criterion in clustering since a stable CH improves the performance of the entire clustered network. CH selection is based on the priority criteria. CH changes should be kept minimum in a network to establish a reliable and stable network. Here, two parameters are taken for CH selection namely the number of followers and average relative velocity of nodes. Nodes connect to the CH by checking the link stability between them. Since VANETs are highly dynamic, checking the average velocity of a particular cluster is important before joining that cluster. When a CH receives a request from a node it calculates its association lifetime with that node. If the association lifetime of a vehicular node is larger than the specified threshold value, only then CH accepts that particular node as a member. The clustering technique provides stability in the network and avoids coordinated sybil attacks.

B. LONGEST COMMON SUBSEQUENCE (LCSS) ALGORITHM

In Multi-channel based sybil attack detection technique, DTW technique is used to find similarity [21]. This is time

consuming since each data point in a sequence is compared with the data points in other sequence. This effects the performance of the detection in real road scenario. In order to overcome this issue, LCSS based dynamic algorithm is proposed for similarity calculation. This technique improves the computation time significantly, thus attackers can be detected in a short duration of time. LCSS is one of the fastest dynamic programming algorithms to measure the similarity between two-time series. It is widely applied in various applications in sensor networks, video and audio processing. Any data which can be represented in a linear sequence can be analyzed by LCSS. The other approaches to model similarity calculation are based on Euclidian distance and DTW distance. Both techniques are relatively sensitive to noise. But LCSS is not sensitive to noise. Few factors need to be considered when distance-based approaches are used in similarity calculations. Those factors are discussed below:

1. Nonuniform Sampling Rates or different speeds: In real time scenarios, the data time series that nodes collect does not guarantee uniformly sampled data. The data collected by sensors in nodes are inconsistent. The time series which is moving in similar fashion but at different speeds will results in large Euclidian distance.

2. **Outliers:** Outliers can be introduced in the data because of anomalies. Even though this may happen at few data points, Euclidian and DTW are insensitive to these outliers. In a security aware topology outliers cannot be avoided at any point [21]

3. **Computation Complexity:** If the data points are more, the computation complexity of normal DTW is O(nm) where n and m are the number of data points. Compared to DTW, LCSS considered in proposed algorithm has complexity of O(n+m).

4. **Non uniform lengths:** Euclidian distance method deals with data sequences of same length. This cannot be possible in a VANET scenario. Here the sequences can be of different lengths because of packet loss due to collision.

Considering all the above challenges, the best suited algorithm for VANET is LCSS. LCSS measures the closeness between two time series to find the maximum number of identical points. This is called Longest Common SubSequence. The LCSS method is explained here in Algorithm 1. A dataset is defined as a collection of values or numbers which are related to an entity. The formula for computing the length of LCSS between two sequences A and B with discrete values is recursive. Consider two datasets X[1...m] and Y[1...n]. A subsequence S[1....s] of X[1..m] can be obtained by deleting m-s data points from X. A common subsequence of X[1...m] and Y[1...n] is a sub-sequence which occurs in both sequences. The longest common subsequence is the subsequence of maximal length. The traditional technique to solve LCSS is to find LCSS of all possible combinations of input time series. If A and B are two input dataset, then the length of LCSS between A and B is dependent on the length of LCSS between the tail of A and B in the following way.

 $LCSS = 0 \text{ if } A = B \tag{3}$

LCSS (tail(A), tail(B) + 1) if A1 = B1 (4)

 $\max\{LCSS(tail(A), B), LCSS(A, tail(B))\}$ if $A1 \neq B1$ (5)

where tail(V)= $\{V2, V3, V4, \dots, V_m\}$

A constant w, called sliding widow control parameter is defined to limit the matching space of dataset A element in dataset B which saves both time and space. It avoids unnecessary comparisons. To compute the similarity, a similarity index, SI is defined between two-time series A and B [22]. SI gives the similarity measure.

$$SI = \frac{Length(LCSS)}{\min(length of A), (length (B))}$$
(6)

$$SI = \begin{cases} 1 \text{ for most similar time series} \\ 0 \text{ for least similar time series} \end{cases}$$
(7)

In recursive algorithm 1 the time complexity is exponential as the intermediate values are computed more than once. Since in VANET the length of the two RSSI sequences are not equal a sliding window is introduced. Once the comparison is done between the sequences as per the sliding window, the rest of the computation for that comparison can be stopped and thus can save time.

Algorithm 1

8
1: function lcss (A, I, m, B, j, n)
2: if $i \ge m$ or $j \ge n$ then
3: return 0
4: else if A[i]=B[i] then
5: return 1+lcss (A, i+1, m, B, j, n)
6: $s1 \leftarrow lcss (A, i+1, m, B, j, n)$
7: $s2 \leftarrow lcss (A, i, m, B, j+1, n)$
8: if $s_1 > s_2$ then
9: return s1
10: else
11. return s2

C. CHANGE POINT ANALYSIS STRATEGY

A changepoint is defined as a sample or time instant at which some statistical property of a data series changes abruptly. The property can be the mean, variance, or a spectral characteristic of the signal. Change Point Analysis (CPA) is an important area of time series analysis [23]. A CPA is performed on a series of time ordered data in order to detect occurrences of changes in the data series. It determines the number of changes and the time of each change. CPA detects abrupt shifts in mean in the time series trends. It can identify anomalous sequences or states in the time series. The main aim of using CPA in proposed algorithm is to detect abrupt mean change in a time series due to power variations imposed by sybil attackers. CPA is incorporated in non-clustered networks in previous works. However, the proposed work combining CPA in clustered networks shows better results in calculating power variations.

The Change Point Detection (CPD) divides each time series into segments, where the mean value within each segment is calculated. The mean value is constant with each segment and changes to a new value at each change point. The goal of CPD is to find the time steps when the mean or standard deviation of the data changes from one value to another [24]. In the new method proposed by us a mean change is calculated and a t-test is performed to find the significant difference between the means of RSSI values from a node. If malicious nodes conduct power control, then similarity comparison fails because when Pt changes RSSI also varies. Hence change point detection using t-test is incorporated [25], [26], [27]. There will be some abrupt changes in the mean values of RSSI time series received from sybil nodes during a detection period. To find the significant difference between the mean values in a time series: T test is performed where t is considered the t-test value, x1 and x2 are the means of the two segments being compared, s1, s2 is the standard deviation of the two segments, and n1 and n2 are the number of observations in each of the segments. T value can be calculated using equation (8).

$$t = \frac{x_1 - x_2}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$
(8)

- t>1: 2 data sets are significantly different.
- t<1: 2 data sets are more similar.

If t>1, the RSSI series from a node undergoes deliberate power control by sybil nodes. It can be detected as sybil attackers.

D. PROPOSED ATTACK DETECTION METHOD

In this section a sybil detection method based on LCSS similarity calculation and change point detection of RSSI time series is addressed. In this method CH and Cluster Members (CM) nodes performs the detection method locally and independently. The nodes need not have any prior trust relationship among themselves. Once the CH detects the illegitimate nodes it transmits the node id of these attackers to all its members as well as the neighbour CHs. This information helps other nodes to select the legitimate nodes as neighbors and avoid attacker nodes. If a multihop connection exists, then CMs detect the sybil nodes in multihop connection. Once the cluster is formed the nodes can communicate to each other only through CHs. This avoids coordination of nodes to launch any kind of attacks. Two sybil attack models considered for the work are:

- 1. Sybil attack launched without power control.
- 2. Sybil attack launched with power control.

The method is performed in 3 stages, acquisition, estimation and verification phase.

1) ACQUISITION PHASE

Safety message frequency is defined as 10Hz in DSRC standard [8]. Each node broadcasts safety message with frequency of 10Hz in common control channel. The CH collects each packet and calculates RSSI of received packet. In a multihop environment parent node also can perform this task. In order to form the time series of RSSI, each node collects RSSI from a particular node for a pre-determined time period called acquisition time. In order to frame the time series CH node stores only the node id and RSSI values [30]. At the end of each acquisition period which is considered as 10 seconds here, a RSSI time series is collected from each node. During this collection period a complete RSSI time series from a node 'i' is received by CH [28], [29].

2) ESTIMATION PHASE

A node performs two techniques to detect sybil attacks in this phase. One method is used to find the attackers which exhibits power control. The other method is used to find the similarity of RSSI nodes to detect the sybil attackers without power control. Once the complete RSSI time series is obtained from a node, CH executes LCCS algorithm and calculates the similarity index as discussed in previous section. It compares the RSSI time sequence of two nodes obtained from a collection period of 10seconds. Each node can transmit 100 packets during this time. Based on the similarity index obtained, the nodes are considered as sybil attackers if they exhibit similar patterns of RSSI values. LCSS based similarity measurement for RSSI received on control interval effectively identifies sybil nodes without power control. But if the malicious node performs power control to its sybil nodes then this method fails. In order to overcome this issue, a t-test is performed for the RSSI series received from each node. RSSI timeseries is a sequence of observations collected from a node. Each time series have different range of values, hence a dataset must be normalised before the analysis for similarity calculations and other processing. Min-Max normalization is the simplest method among many other methods for normalization [31], [32]. This method helps us to understand the data easily. The RSSI values are normalised using the equation (9).

$$RSSI_{i} = \frac{RSSI - RSSI_{min}}{RSSI_{max} - RSSI_{min}}$$
(9)

This method arranges most of the values between [0,1]. The outliers are taken care by this calculation. RSSI _{min} and RSSI_{max} are the minimum and maximum values of RSSI in the time series.

3) VERIFICATION PHASE

During acquisition phase each node can get the number of change points and LCSS distances for all neighbouring nodes. LCSS distance between all sybil nodes fabricated by attacker should be very small close to 1. But distance between sybil nodes and normal nodes are close to 0.



FIGURE 2. Block diagram of sybil detection.

The complete block diagram is shown in Figure 2. The nodes receive the RSSI time series and stores it with the node ids. The method implements two techniques to find the malicious nodes. LCSS based similarity comparison technique is a good choice if the nodes are not performing power control. In the first technique similarity comparison is performed by the nodes between its neighbours. Once the observation period is completed the nodes performs similarity comparison. Similarity Index between the normal nodes and between normal and sybil nodes have low index as their distances are different in real. If the similarity index is greater than threshold value, the nodes are detected as sybil. However, if the nodes perform power control, then the first method cannot detect attackers. So, second technique is performed. In this technique once the RSSI from each node are received the mean change point detection is performed. The cumulative average of segment means is calculated. This is taken as the threshold value. Then the mean of each segment is compared with the threshold value. If the segment mean is greater than the threshold value then a change point is detected. After calculating the average number of change points, a node can be considered as an attacker. If the number of change points are more than 50% of the total segments, the nodes are considered as sybil nodes. However, the density has a higher impact on the detection method. Suppose if a node is in a signal junction or in a traffic jam then the distance from the nodes becomes closer. Hence, it will be very difficult to distinguish between the normal and sybil nodes. The RSSI received from sybil and normal nodes will become similar in these cases. Another issue during the above conditions is the number of packets lost due to collision also increases. Since CH is running this algorithm and when the attacker gets detected, it can transmit the information to all members at a faster rate.

IV. PERFORMANCE EVALUATION AND RESULTS

In this section, the performance of proposed technique LCSS based Sybil Detection in Clustering Network(LSDCN)

is compared with Two-Layer Intrusion Detection Technique(TIDS) and LCSS based Sybil Detection in Unclustered Network (LSDUN) is compared with multichannel Voiceprint technique [11], [4]. NS3 (release 3.34) network simulator is used for simulating the results. The vehicle traces obtained from MOVE-SUMO traffic simulator are used to feed traffic scenarios for NS3 simulation.

The simulation parameters used are as shown in Table 2.

TABLE 2. Simulation parameters.

Parameters	Values
Simulation time	300s
Road length/topology	3km, 2-way highway
Road Segment	1km
Max Speed	10-35m/s
No of vehicles	50-100
Transmission range	100-300m
Propagation Model	Shadowing

The simulation is carried out for 300s and multiple runs are performed to obtain better results. The performance of the algorithm is evaluated using detection rate or true positive rate, false positive rate and detection delay. The results are obtained for various speeds from 10m/s (36km/h) to 35m/s (144km/h) and the transmission range between 100m to 300m based on DSRC standard. 5-15% vehicles are considered malicious in the network. The beacon size is 100 bytes maximum. The proposed algorithm improves the network performance parameters like detection rate and detection delay in the range between 8% to 10% compared to Twolayer collaborative Intrusion Detection System(TIDS) [11]. The comparison between these algorithms is explained in the results.

A. DETECTION RATE

Detection rate (DR) or true positive is the ratio of correctly detected sybil nodes to total illegitimate nodes. Figure 3(a) shows the comparison of detection rate with vehicle density in two similar unclustered networks. As the vehicle density increases detection rate shows a reducing tendency. All methods do not require any cooperative detection method where a node needs the RSSI values observed by its neighbouring nodes for attack detection. Both methods do not require any trust relationship among nodes. Voice print and unclustered network has similar detection rate as in Figure 3(a). Voiceprint is sensitive to outliers, hence RSSI data points with significant differences are not considered during data acquisition. But proposed method, LSDUN is insensitive to outliers. Also, each CH is performing the attack detection that enable the network to detect more attackers



FIGURE 3. Detection rate vs vehicle density.

in a shorter period. As density increases, there are two possibilities for reduction in detection rate. At high densities, the number of vehicles is more in the network and the packets exchanged between them also increases proportionally. This results in information loss due to collision of packets. So, the RSSI series received by the CH becomes less, which leads to less accuracy in detection. Also, when the number of vehicles is more the relative distance between the vehicles decreases.

The RSSI calculation is mainly based on two parameters, distance and transmitted power. For a constant transmitted power if the distance is relatively small the RSSI series received from normal nodes and malicious nodes shows high similarity. Due to this reason, it is difficult to identify the malicious nodes from normal nodes. So, the detection rate decreases. However, the proposed method can detect all the attackers using LCSS based similarity technique. Hence the detection rate shows an improvement of 5% at higher density compared to Voiceprint. Data rate is calculated using equation 10.

Detection rate =
$$\frac{no \ of \ Fabricated \ nodes}{malicious + sybil \ nodes}$$
 (10)

Figure 3(b) shows a DR comparison of collaborative detection methods using Two-layer collaborative IDS [11] with LSDCN. In this method, RSU and nodes determine the attacker. A machine learning based approach is performed in the nodes and RSUs. The final decision is taken by the RSU. In a dynamic environment, independent detection by the nodes is better compared to infrastructure based approach. The nodes need to be in the range of RSU and RSU is static. Since the proposed approach does not depend on any infrastructure and the decisions are done by the moving CH, the detection rate has slight improvement compared to TIDS. As the density increases 100% reception of RSSI value is not possible due to collisions. Hence there is a reduction in detection rate.

B. FALSE POSITIVE RATE

False positive rate (FPR) is the rate of normal nodes incorrectly identified as malicious nodes. FPR is calculated using equation 11. Figure 4(a) shows the false positive rate vehicle density of LSDUN and Voiceprint. The distance between the nodes will be relatively high when the vehicles are sparse. Hence FPR is less initially. However, the proposed method and Voiceprint is able to detect the false positives since both incorporates the dynamic method to calculate optimum decision boundary. The reduction in average distance results in nodes being falsely identified as illegitimate nodes. As density increases due to the collision of packets, packet drop increases. This further increases FPR. False Positive Rate is minimal till the vehicle density reaches 30 vehicles/km in both cases. Compared to other existing technique, since the density and LCSS distance are considered to calculate the optimal boundary, FPR is reduced in the proposed method. However, as the density increases the distance between the nodes reduces marginally and there is an increase in FPR.

Figure 4(b) shows a FPR comparison of collaborative detection methods using Two-layer approach(TIDS) and LSDCN. In TIDS the On Board Unit(OBU) and Road Side Unit(RSU) performs the detection using random forest classifiers. It is a complex algorithm and all nodes execute the first tier detection and there is a possibility that the information exchanged between the nodes and RSU may undergo packet loss due to collisions. In clustered networks this issue is relatively less because CH receives RSSI from its members only. The FPR is low in the proposed method since distance is calculated as a function of density. TIDS does not consider this issue at all and this increases the FPR as the vehicle density increases. An average improvement of 8-10%



FIGURE 4. False positive rate vs vehicle density.

is obtained by the proposed system.

$$FPR = \frac{no \ of \ wrongly \ identified \ legitimate \ nodes}{Total \ normal \ nodes}$$
(11)

C. DETECTION DELAY

The detection delay is the time difference between when the attacker enters the network and when they get detected. As the number of nodes increases the number of samples received also increases. So, there is a linear increase in detection delay. Detection delay is mainly based on the computation time of the algorithm. Figure 5(a) shows a linear increase in computation time as samples increase in both Voiceprint and unclustered networks when the number of attackers are increasing. As the number of attackers are increasing the RSSI received by the node in LSDUN and Voiceprint increases. Voiceprint is based on DTW which requires high

computation time in RSSI similarity calculation compared to LCSS. Hence the detection delay increases.



FIGURE 5. Average detection delay vs no. of attackers.

But the proposed method incorporates LCSS techniques and the detection is very fast as the number of attackers increases. In a clustered network LCSS algorithm takes less time as it reduces the number of comparisons which reduces the computation time of the algorithm as shown in Figure 5(b). Also since cluster head takes decisions based on the detection algorithm it will receive RSSI values from its members only. Compared to TIDS, LCDCN shows an improvement of 15% in computation time, thus improving the detection delay. TIDS is a complex machine learning algorithm that is implemented in RSU because of the resource constraints of OBUs. But the proposed method is a lightweight algorithm that can be used by any node. A significant improvement is shown by the proposed method.

D. OVERHEAD

A comparison graph between the overhead of LSDCN and TISB is shown in Figure 6. The LSDCN implements the clustering technique. Hence, clustering overhead is contributing to the total overhead of the network. During cluster formation cluster heads and the members exchange information. When the cluster heads perform cluster maintenance few message exchanges happen.



FIGURE 6. Overhead vs velocity.

All these handshaking contribute to overhead. However, since the handshaking messages between the CH and the members are minimal in MPMCA, the overall overhead is very less compared to other method. In TISB a machine learning-based two tier detection strategy is employed in OBU and RSU. The number of messages getting exchanged is more here. Since it is an RSU based detection method all the information from all nodes has to reach the RSU. But in the proposed method since the cluster head is performing the detection techniques, the main contributor to overhead is the clustering overhead only. In MPMCA it is observed that for a transmission range of 300m, the average overhead is less than 10%. The similarity calculation adds a slight overhead in the proposed method.

E. LCSS SIMILARITY INDEX AND AVERAGE NUMBER OF CHANGE POINTS

Figure 7 shows similarity index in terms of LCSS distance for RSSI comparison in clustered network. Similarity index calculation in clustered network were not implemented in earlier works to best of our knowledge and this method contributed an improved result in the proposed method. Nodes transmit with constant power during entire observation period. If the attackers do not perform power control, then



FIGURE 7. LCSS distance vs detection time.

RSSI values between the normal nodes and the sybil nodes shows significant variations. Then the LCSS distance will be less and the similarity index also will be closely to 0. At the same time since the RSSI values between the sybil nodes and malicious nodes are similar because of distance factor, the similarity index will be high. It is observed that the similarity index among the sybil nodes and the malicious nodes are close to 1. A threshold is calculated in this case by taking the average of all SI from all neighbouring nodes by the detecting node. By taking multiple detection times the value is fixed at 70% in the simulation. If the SI is greater than 70% then those nodes are considered as sybil.



FIGURE 8. No. of change points, (N_{CP}) vs detection times.

Nodes can be identified easily using RSSI comparison if there are no power variations happening in the network. But if the attackers deliberately perform power control then a changepoint detection technique based on mean values is performed. In the Figure 8 it can be clearly observed that the



FIGURE 9. Comparison of detection rate.



FIGURE 10. Comparison of false positive rate.



FIGURE 11. Comparison of average detection delay.

RSSI values of normal nodes does not vary significantly. The number of change points detected for normal nodes will be very less during detection times. But malicious node performs power control to sybil nodes. So, there will be high variations in the RSSI values received from those nodes. The average number of change points observed during multiple detection times is more in sybil nodes.

The comparison of the performance parameters of the proposed work with existing algorithms [11], [4] is shown in Figure 9, Figure 10 and Figure 11.

V. CONCLUSION

Sybil node detection is a major research area in wireless vehicular network. In this manuscript, a novel sybil attack detection mechanism using a novel hybridization technique is proposed. The method incorporates LCSS and CPD techniques which can detect sybil attackers with and without power control in clustered and unclustered networks. The computation complexity of LCSS based algorithm is very minimal which aids the time delay requirements of VANET. This method is extended by incorporating a mean value-based change point detection method to check any abrupt changes occurring in RSSI time series. Any kind of coordination between the nodes are not be possible since the algorithms are performed in a controlled clustered environment by the cluster head. This helps in avoiding the coordinated attacks in the network. Proposed method neither require any support from RSU nor any trust-based relationship among neighbours. The overall results are compared with a existing techniques. The results show a significant improvement in terms of detection delay and detection rate.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript.

REFERENCES

- Z. A. Abdulkader, A. Abdullah, M. T. Abdullah, and Z. A. Zukarnain, "A survey on Sybil attack detection in vehicular ad hoc networks (VANET)," *J. Comput.*, vol. 29, no. 2, pp. 1–6, 2018.
- [2] R. Hussain and H. Oh, "On secure and privacy-aware Sybil attack detection in vehicular communications," Wireless Pers. Commun., vol. 77, no. 4, pp. 2649–2673, Aug. 2014, doi: 10.1007/s11277-014-1659-5.
- [3] Y. Yao, B. Xiao, G. Yang, Y. Hu, L. Wang, and X. Zhou, "Power control identification: A novel Sybil attack detection scheme in VANETs using RSSI," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 11, pp. 2588–2602, Nov. 2019, doi: 10.1109/JSAC.2019.2933888.
- [4] Y. Yao, B. Xiao, G. Wu, X. Liu, Z. Yu, K. Zhang, and X. Zhou, "Multi-channel based Sybil attack detection in vehicular ad hoc networks using RSSI," *IEEE Trans. Mobile Comput.*, vol. 18, no. 2, pp. 362–375, Feb. 2019, doi: 10.1109/TMC.2018.2833849.
- [5] R. Douceur, "The Sybil attack," in Proc. 1st Int. Workshop Peer-to-Peer Syst., 2002, pp. 251–260.
- [6] M. Demirbas and Y. Song, "An RSSI-based scheme for Sybil attack detection in wireless sensor networks," in *Proc. Int. Symp. World Wireless*, *Mobile Multimedia Netw. (WoWMoM)*, 2006, p. 5, doi: 10.1109/WOW-MOM.2006.27.
- [7] M. Baza, M. Nabil, M. M. E. A. Mahmoud, N. Bewermeier, K. Fidan, W. Alasmary, and M. Abdallah, "Detecting Sybil attacks using proofs of work and location in VANETs," *IEEE Trans. Depend. Sec. Comput.*, vol. 19, no. 1, pp. 39–53, Jan./Feb. 2022.
- [8] M. T. Garip, P. H. Kim, P. Reiher, and M. Gerla, "INTERLOC: An interference-aware RSSI-based localization and Sybil attack detection mechanism for vehicular ad hoc networks," in *Proc. 14th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2017, pp. 1–6.

- H. B. Tulay and C. E. Koksal, "Robust Sybil attack detection in vehicular networks," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2021, pp. 1–7.
- [10] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2418–2434, Jun. 2010, doi: 10.1109/TVT.2010.2044904.
- [11] P. H. Mirzaee, M. Shojafar, H. Bagheri, T. H. Chan, H. Cruickshank, and R. Tafazolli, "A two-layer collaborative vehicle-edge intrusion detection system for vehicular communications," in *Proc. IEEE 94th Veh. Technol. Conf. (VTC-Fall)*, Norman, OK, USA, Sep. 2021, pp. 1–6, doi: 10.1109/VTC2021-Fall52928.2021.9625388.
- [12] D. Zhang, H. Ge, T. Zhang, Y.-Y. Cui, X. Liu, and G. Mao, "New multi-hop clustering algorithm for vehicular ad hoc networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 4, pp. 1517–1530, Apr. 2019, doi: 10.1109/TITS.2018.2853165.
- [13] O. Senouci, Z. Aliouat, and S. Harous, "MCA-V2I: A multi-hop clustering approach over vehicle-to-internet communication for improving VANETs performances," *J. Future Gener. Comput. Syst.*, vol. 96, pp. 309–323, Jul. 2019.
- [14] D. Zhang, C. Gong, T. Zhang, J. Zhang, and M. Piao, "A new algorithm of clustering AODV based on edge computing strategy in IOV," *Wireless Netw.*, vol. 27, no. 4, pp. 2891–2908, May 2021.
- [15] A. Alsarhan, Y. Kilani, A. Al-Dubai, A. Y. Zomaya, and A. Hussain, "Novel fuzzy and game theory based clustering and decision making for VANETs," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1568–1581, Feb. 2020.
- [16] X. Cheng and B. Huang, "A center-based secure and stable clustering algorithm for VANETs on highways," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–10, Jan. 2019.
- [17] G. H. Alsuhli, A. Khattab, and Y. A. Fahmy, "Double-head clustering for resilient VANETs," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–17, Mar. 2019.
- [18] M. Fathian and A. R. Jafarian-Moghaddam, "New clustering algorithms for vehicular ad-hoc network in a highway communication environment," *Wireless Netw.*, vol. 21, no. 8, pp. 2765–2780, Nov. 2015.
- [19] L. Bergroth, H. Hakonen, and T. Raita, "A survey of longest common subsequence algorithms," in *Proc. 7th Int. Symp. String Process. Inf. Retr.* (SPIRE), 2000, pp. 39–48, doi: 10.1109/SPIRE.2000.878178.
- [20] R. Khan, I. Ali, S. M. Altowaijri, M. Zakarya, A. U. Rahman, I. Ahmedy, A. Khan, and A. Gani, "LCSS-based algorithm for computing multivariate data set similarity: A case study of real-time WSN data," *Sensors*, vol. 19, no. 1, p. 166, Jan. 2019.
- [21] D. Cao and J. Liu, "Research on dynamic time warping multivariate time series similarity matching based on shape feature and inclination angle," *J. Cloud Comput.*, vol. 5, no. 1, Dec. 2016, Art. no. 11, doi: 10.1186/s13677-016-0062-z.
- [22] R. Mariescu-Istodor and P. Fränti, "Context-aware similarity of GPS trajectories," J. Location Based Services, vol. 14, no. 4, pp. 231–251, Oct. 2020, doi: 10.1080/17489725.2020.1842923.
- [23] W. A. Taylor, "Change-point analysis: A powerful new tool for detecting changes," Taylor Enterprises, Spartanburg, SC, USA, Tech. Rep., 2000, pp. 1–19. [Online]. Available: https://variation.com
- [24] S. Deldari, D. V. Smith, H. Xue, and F. D. Salim, "Time series change point detection with self-supervised contrastive predictive coding," in *Proc. Web Conf.*, Apr. 2021, pp. 3124–3135, doi: 10.1145/3442381.3449903.
- [25] C. Barrado, M. Pérez-Batlle, M. Lopez, and E. Pastor, "Paired T-test analysis to measure the efficiency impact of a flying RPAS in the nonsegregated airspace," in *Proc. IEEE/AIAA 35th Digit. Avionics Syst. Conf.* (*DASC*), Sep. 2016, pp. 1–7, doi: 10.1109/DASC.2016.7778008.
- [26] T. Riaji, S. E. Hassani, and F. E. M. Alaoui, "Application of paired samples t-test in engineering service-learning project," in *Proc. 11th Int. Symp. Signal, Image, Video Commun. (ISIVC)*, 2022, pp. 1–4, doi: 10.1109/ISIVC54825.2022.9800209.
- [27] A. Gupta, P. Mishra, C. Pandey, U. Singh, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," *Ann. Cardiac Anaesthesia*, vol. 22, no. 1, pp. 67–72, 2019.
- [28] S. Kanumalli, A. Ch, and P. Murty, "An efficient method for detection of Sybil attackers in IOV," *Int. J. Adv. Model. Anal. B*, vol. 61, no. 1, pp. 5–8, Mar. 2018.
- [29] M. Jamshidi, A. M. Darwesh, A. Lorenc, M. Ranjbari, and M. R. Meybodi, "A precise algorithm for detecting malicious Sybil nodes in mobile wireless sensor networks," *IEIE Trans. Smart Process. Comput.*, vol. 7, no. 6, pp. 457–466, Dec. 2018, doi: 10.5573/IEIESPC.2018.7.6.457.

- [30] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowl. Inf. Syst.*, vol. 51, no. 2, pp. 339–367, May 2017, doi: 10.1007/s10115-016-0987-z.
- [31] H. W. Herwanto, A. N. Handayani, A. P. Wibawa, K. L. Chandrika, and K. Arai, "Comparison of min-max, Z-score and decimal scaling normalization for zoning feature extraction on Javanese character recognition," in *Proc. 7th Int. Conf. Electr., Electron. Inf. Eng. (ICEEIE)*, Oct. 2021, pp. 1–3, doi: 10.1109/ICEEIE52663.2021.9616665.
- [32] P. J. M. Ali, "Investigating the impact of min-max data normalization on the regression performance of K-nearest neighbor with different similarity measurements," *ARO-Sci. J. Koya Univ.*, vol. 10, no. 1, pp. 85–91, 2022. [Online]. Available: https://orcid.org/0000-0002-0471-5172
- [33] O. Senouci, S. Harous, and Z. Aliouat, "A new heuristic clustering algorithm based on RSU for Internet of Vehicles," *Arabian J. Sci. Eng.*, vol. 44, no. 11, pp. 9735–9753, Nov. 2019.
- [34] A. B. Tambawal, R. M. Noor, R. Salleh, C. Chembe, and M. Oche, "Enhanced weight-based clustering algorithm to provide reliable delivery for VANET safety applications," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0214664.
- [35] W. Ahsan, M. F. Khan, F. Aadil, M. Maqsood, S. Ashraf, Y. Nam, and S. Rho, "Optimized node clustering in VANETs by using meta-heuristic algorithms," *Electronics*, vol. 9, no. 3, p. 394, Feb. 2020.
- [36] A. Katiyar, D. Singh, and R. S. Yadav, "State-of-the-art approach to clustering protocols in VANET: A survey," *Wireless Netw.*, vol. 26, no. 7, pp. 5307–5336, Oct. 2020.
- [37] M. Bersali, A. Rachedi, and H. Bouarfa, "A new collaborative clustering approach for the Internet of Vehicles (CCA-IoV)," in *Proc. 2nd Int. Conf. Embedded Distrib. Syst. (EDIS)*, Nov. 2020, pp. 58–63.
- [38] S. Azam, M. Bibi, R. Riaz, S. S. Rizvi, and S. J. Kwon, "Collaborative learning based Sybil attack detection in vehicular AD-HOC networks (VANETS)," *Sensors*, vol. 22, no. 18, p. 6934, Sep. 2022, doi: 10.3390/s22186934.



S. RAKHI (Member, IEEE) received the M.Tech. degree (Hons.) in VLSI and embedded systems from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India, in 2010, where she is currently pursuing the Ph.D. degree. She is also an Assistant Professor with the Department of Electronics and Communication Engineering, Atria Institute of Technology, Bengaluru, Karnataka. Her research interests include security in wireless networks and vehicular networks. She

was a recipient of AWSAR Award for best research article (2019–2020) from the Department of Science and Technology, Government of India. She has presented research papers in the areas of VLSI, vehicular ad-hoc networks, and wireless networks PA, in 2008.



K. R. SHOBHA (Senior Member, IEEE) received the M.E. degree in digital communication engineering from Bengaluru University, Karnataka, India, and the Ph.D. degree from Visvesvaraya Technological University. She is currently a Professor with the Department of Electronics and Telecommunication Engineering, Ramaiah Institute of Technology, Bengaluru. She has more than 25 papers publications to her credit. Her research interests include mobile ad-hoc networks, the IoT,

and cloud computing. She is also serving as the Chair for IEEE Bangalore Section Sensors Council Chapter and an ExCom Member for WIE and IEEE Communication Society, Bengaluru. She is also an active member of IETE, ISoc, and IAENG.