IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 01-07
www.iosrjournals.org

A collaborative IDS for Vehicular Ad-Hoc Networks C-VIDS using Data Mining Technique

Dr S.B.Ninu, *Associate Professor -PG Department of Computer Science, Thiruthangal Nadar College*
Mr.T.Prem Kumar, *Assistant Professor of Computer Science Thiruthangal Nadar College*

Abstract:
Vehicular ad hoc network (VANET) is a subclass of MANETs are vulnerable to various kinds of threats due to their dynamic nature and lack of central point of control. Vehicles (nodes) in VANETs share real-time information about their movements, traffic and road conditions. Existing cooperative IDS are vulnerable that share misleading and manipulated information and disrupts the IDS normal condition. Hence, in this paper proposed an intelligent collaborative model based on data mining for intrusion traffic detection system that can detect the attacks. As such, find friendly vehicle nodes in the network that continuously monitor the behavior of other vehicle nodes to find the anomalous behavior. For the performance of the proposed system NS-2 simulations were carried out. To evaluate the performance of proposed collaborative IDS scheme the various existing IDS models are used. The results clearly states that the proposed IDS considerably reduces the false positive rate, thereby proving that the proposed technique is capable of identifying anomalies in network better than other existing system.
Keywords: *VANET, Anomaly detection, Data mining, IDS, Collaborative IDS*

I. Introduction:

Ever increasing demand of the digital era has forced the researchers to continuously develop new emerging trends, particularly the wireless communication services, However security threats in the VANET present considerable challenges.Vanets being an Adhoc-network are at risk of misbehaviors because of lack of centralized administration [1].The malicious vehicular nodes can join the network and launch various types of attacks due to the lack of central point of control [2].The Vanet environment is highly dynamic with rapid dynamic topology in which the vehicle nodes are varying in speeds and density [3].the high mobility, varying density and network size introduce new vulnerabilities and challenges when applying IDS in vanet [4].

Many solutions have been proposed to protect vehicles from being a target of cyber-attacks.Prevention techniques digital signature, authentication, and encryption have been widely used as a first line of defense to prevent many types of external attacks.However, these preventive measures are in adequate for protection against the insider attacks.Due to the cooperative nature of VANET,malicious nodes or intruders can still perform malicious Denial of services [5-8]. An attack can be launched by a single node or multiple nodes in a cooperative manner. In internal and external attackers the internal attackers is the most dangerous and difficult to detect. In some attacks, multiple attackers synchronize their actions to disrupt a target network.

However, due to the cooperativenature of VANET, many of the recent proposed IDSs rely on the collaboration between vehicles todetect the intruders [9].In the cooperative IDS (CIDS), vehicles share knowledge related totheir detection experiences to help vehicles in the vicinity to detect the intruders more accurately. Motivated by this, collaborative IDS have been developed, with the purpose of strengthening a single IDS by collecting knowledge and learning experience from other vehicle IDS nodes. According to [10], collaborative IDS is expected to enhance the overall detection accuracy of intrusion vehicle assessment and will also improve the possibility of identifying novel attacks.

The main objective of this paper is to design a robust collaborative vehicular IDS that can effectively evaluate the trustworthiness of each vehicle node within the network and identify the intrusions in the network.

The collaborative Vanet IDS proposed in this paper uses data mining techniques for detecting attacks. Here we select the friendly vehicle nodes based on their trust that continuously monitor the behavior of other vehicle nodes in the network for any intrusions. The proposed Vanet IDS was evaluated by NS- 2 simulation which showed that the proposed system considerably reduced the false positive rate compared to other existing IDS, thereby proving that it is better than other existing systems.

Related Work:

Securing VANETs has attracted great interest of many researchers during the last years [11-18].VANET is vulnerable to many security issues that can disrupt the functionality of these applications.The intrusion detection system for VANETs aims to detect internal as well as external attacks with high accuracy [17-18].

Acknowledgement scheme for Vanets requires all acknowledgement packets to be digitally signed by its sender and verified by its receiver .They used DSA and RSA as digital signatures and showed that this scheme is able to detect wide range of attacks. The drawback of this scheme is the requirement to digitally sign all the acknowledgements which increases computational overhead.

An watchdog IDS scheme of VANET which consists of two different modules, the watchdog and pathrater.In this scheme, the watchdog acts as an IDS for the VANET and detect malicious node behaviors in the network by listening to its next hop's transmission. If the Watchdog notices that itsimmediate next node fails to forward the packet within agiven period of time then it increments the node's failurecounter.Ifthefailurecounterofthemonitorednodeexceeds a threshold value then the Watchdog reports thenode as misbehaving. The Pathrater is then employed toinform the routing protocol to avoid the reported nodes forfurther data transmission. The drawback of this scheme isthat it requires continuous monitoring by the Watchdog fordetectingintrusions.

MachinelearningmethodswereappliedwidelytosolveIDSissuesindifferentnetworks.For instance, the random forest (RF) method [13] to build automaticallypatterns of intrusions.Then, intrusions were detected by matching the network activities againstthe built patterns.To evaluate the performance of this model, the authors used the knowledgediscovery data mining (KDD)'99 dataset.To handle the problem of imbalanced data, the two sampling down and over methodswere applied. After that, the simulation was carried out onthe WEKA environment for training and testing set. The experimentalresults showed that this approach achieved a high-detection rate of 94.7% with a low false-positiverate of 2%.In [19], the combination of the k-nearest neighbor(K-nn)methodwithageneticalgorithmwasused.

A hybrid IDS model for Wireless Local Area Network (WLAN)that uses both misuse and anomaly based IDS sub-modules to detect intrusion.The drawback of this approach is that the response times of the misuse based and anomaly based IDSs are different. It also introduces significant computational overhead due to processing of the same data traffic by two different IDSs.

Recently, several research have been published related to Machine Learning for intrusion detection in VANET.For instance, Shams et al. [20] proposed an approach combining the promiscuous mode for data collection and SVM for IDS in VANET.They aimed to analyze data to create a trust value for vehicles on the network as trust aware SVM-based IDS. The main idea was to guarantee that vehicles withinthe network have a complete idea about activities of their next hop, which will help to maintain high-performance in case of attacks.

In the cooperative IDS (CIDS), vehicles share knowledge related totheir detection experiences to help vehicles in the vicinity to detect the intruders more accurately.

II. Proposed Methodology:

This paper propose a framework for detecting intrusion in VANET with the help of other specific nodes in thenetwork. The proposed collaborative vehicle C-VIDS work in collaboration withother designated IDS vehicle nodes in the network. Mostly the friendly nodes are designated as IDS nodes. All the IDS nodes continuously monitor the behavior of other vehicle nodes inthe network.

The nodes used in the C-VIDS maintain a list for formulating the attacks, and further, it has the information regarding the other friendly nodes in the VANET. The IDS node established in the MANET is referred to as the helpmate node. The IDS nodes employed in the framework have the freedom to choose its own friendly nodes. As provided in the architecture, the scheme employs the trust model for detecting the intruder nodes. The trust model has the information regarding the feedback node, and the packet information. The entire model for the intrusion detection is explained as follows:

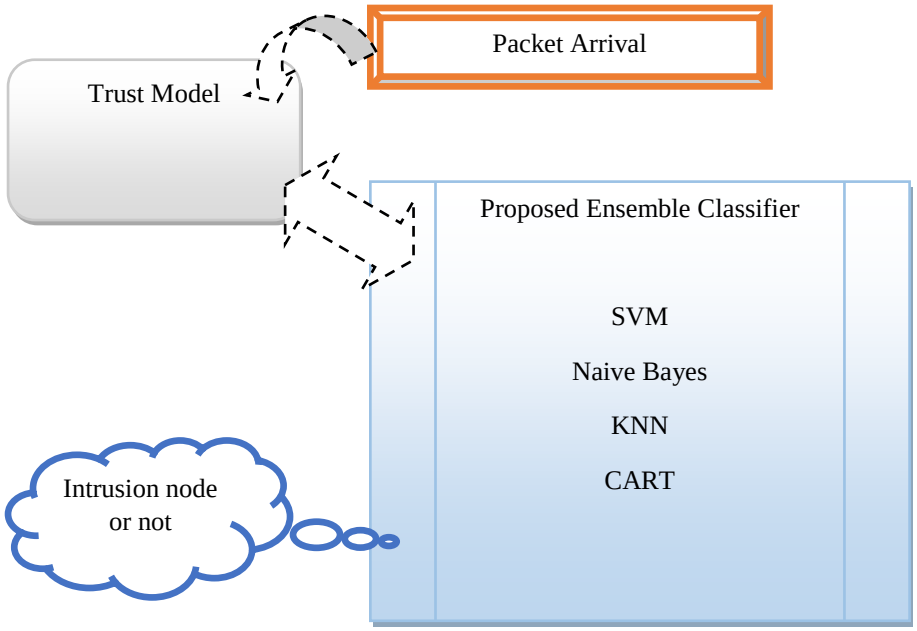

Figure 1 Architecture of the Proposed Collaborative Data Mining Algorithm for C-VIDS

Each IDS node maintains a list called *helpmates*that havedetails of otherfriendly IDSnodeswithwhich itcurrently collaborates with. Each IDS node in the networkhas the freedom to choose its *friends* based on their owninterest. The communications between collaborating IDSnodes are requests for intrusion alert evaluation and their corresponding feedbacks. An IDS node initiates a cooperative intrusion detection procedure under any of the two conditions. The first condition is that the IDS node cannot detect locally an intrusion anomaly and the second condition is that the evidence available locally is in conclusive and warrants broader investigation. The overall procedure works by propagating the intrusion detection state information among neighboring IDS nodes.

The helpmates are found by their trustworthiness. To evaluate the trustworthiness of a friendly IDS node, anIDS node can send a challenge to this target periodically using a random generation process (i.e., sendingtimeisnot fixed, but random). When receiving the feedback fromthe target node, the IDS node can give a score to reflect itssatisfactionlevel.Sincewedefinetwotypesoftrustincluding feedback-based trust (T_{fd}) and packet-based trust(T_{pt}), we develop a single metric called overall trust (T_{total})to facilitate the trust evaluation as follows:

$$T_{total}=W_1\times T_{fd}+W_2\times T_{pt} \qquad (1)$$

WhereW_1andW_2areweightvaluesand$W_1+W_2=1$.

For the feedback based trust $T^{i,j}$of node i according tonode j, we can compute it by using the equation describedasbelow:

$$T_{fd}^{i,j} = W_s \frac{\sum_{c=0}^{n} F_c^j \lambda^{ct}}{\sum_{c=0}^{n} \lambda^{ct}} \qquad (2)$$

where, F_c^j indicates the score of the received feedback c and n signifies the total feedbacks provided to the system. The term λ refers to the forgetting factor. The term W_s refers to the significant weight and it has the value of $W_s = \frac{\sum_{c=0}^{n} \lambda^{ct}}{m}$. The next trust factor used by the proposed trust scheme is expressed as,

$$T_{pt}^{i,j} = \frac{c+1}{N+2} \qquad (3)$$

where, c and N signify the total received benign packets and received packets. The trust model incorporates the trustworthiness of a node j based on the weighted majority expressed as,

$$T_j = \frac{\sum_{T>r} T_{total}^{i,j} D_j^i I_S^i}{\sum_{T>r} T_{total}^{i,j} D_j^i} \qquad (4)$$

s where, r signifies the threshold for the alert ranking. The terms I_S^i and D_j^i signify the intrusion sensitivity and the hops amidst the nodes i and j. The trust value of the node j can have value amidst [0,1] and the information is fed to the classifier.

Once the helpmates *are* found the IDS nodes uses asystemthatcontainsanensembleoffourmethodsfordetectingintrusion.Thesemethodsworkasfollows:

a) Supportvectormachine(SVM)

SVM[20] is a supervised machine learning method for classification and regression analysis.The principle of SVM is to find the best hyper plane to separate the data into two parts. A SVM model consists of the samples represented as points in space. The samples ofthe different categories are divided by hyper plane. This hyper plane always maximizes the margin between those two regions or classes. The margin is defined by the farthest distance between the samples of the two classesand computed based on the distances between the closest samples of both classes, which are called supporting vectors.
Testsamplesarethenmappedintothesamespace. Based on whichsidesof the hyper plane they fallon,testsamplesarepredictedtobelongtothecorrespondingcategories.

b) Classificationandregressiontree (CART)

CART is a method based on the Giniindex.It usually uses a top down approach when CART constructs a decision tree. Decision tree [21] is a categorization model that recursively partitions the training data into a tree structure. In the experiments, we first put all the training samples at the root node. We then search the best partition of the root node so that the Giniimpurity can reduce to minimum. Gini impurity represents the possibility that a randomly selected sample is classified into the wrong subset.When all the samples of a node belong to one class,Giniimpurity equals to zero.We use the best partition to divide root node into two parts, each of which is seen as a new node.This process is then repeated on the new nodes.

c) NaiveBayes(NB)

Naïve Bayes is a probabilistic classifier based on Bayes theorem.Given test sample,we need to calculate the probabilities of the appearance of various categories under the condition of the appearance of a test sample.The sample belongs to the category whose probability is the largest.

d) K-NearestNeighbor(K-NN)

K-NN algorithm is an on-parametric statistical methods for categorization and regression. It classifies a test sample by measuring the distance between the training samples and test sample. We need to choose k nearest samples and use majority voting to predict which category the sample belongs to.

To exert the advantage of each algorithm and to further improve the accuracy of detectionand categorization,we employ the ensemble of multiple classifiers with majority voting after obtaining the classification results of the five algorithms described above.When a testapp is given as input, each base classifier predictsits classification.All the five prediction results will then vote to generate a final prediction.

Ensemble Classification

After providing the trust model to the ensemble classifier, the majority voting is enabled by the proposed scheme. The proposed scheme analyzes the results of the four classifiers and votes the result having the major impact on the detection result. As the ensemble classifier enables the advantages of the classifier, the classification results are more accurate.

III. Results and Discussion:

We have implemented our proposed model C-VIDS in thenetwork simulator NS2 installed on Ubuntu 12.04 runninggcc version 4.6.3. We restrict the movements of mobile nodes to a predefined flat grid area.Table-1lists the various parameters used for our simulation.

Parameters	Value
Simulation	8000-15000s
NumberofNodes	50,100,150
SimulationArea	600x600 m2
TransmissionRange	150m
Mobility	RandomWayPoint
RoutingProtocol	DSR
MACLayer	DCFofIEEE802.11
Max.NodeMovementSpeed	20m/s
PauseTime	500s
TrafficType	CBR/UDP
PacketSize	512Bytes
MAC Protocol	IEEE 802.11
Sampling Interval	3s

Table-1.NS2simulation-Parameters.

Evaluation Metrics

The performance analysis of the proposed collaborative data mining algorithm is done through six metrics, classification accuracy, TPR, FPR, F-measure, PDR, and routing overhead.

TCL scripts were used to generate which containdata related to normal profile and simulated attacks. Thesampling rate of 3 s is used to record the values. The rules extracted from the set races are then used to build the normal and abnormal behaviour of the network. In order tocompare efficiency of proposed algorithm, the test traceswere used to compute the classification accuracy using the following parameters:

- Accuracy
- TruePositiveRate(TPR)
- FalsePositiveRate(FPR)
- Precision
- Recall
- F-measure

The detection results of the proposed IDS usingeachclassifierareshowninTable-2.Itisseenthatthe TPR of ensemble are the highest among the five methods,achieving 98.25%. This method also achieves the accuracyof 99.39% in the detection of intrusion after employing ensemble of the four classifiers with majority voting mechanism.In general,our method outperforms SVM,CART,NB and K-NN.

Classifier	**Accuracy** (%)	**TPR** (%)	**FPR** (%)	**Precision** (%)	**Recall** (%)	**F-measure** (%)
SVM	98.82	96.07	3.39	92.79	95.09	93.93
CART	99.23	95.83	0.52	93.31	95.83	94.55
NB	76.46	90.92	24.63	21.73	90.92	35.08
K-NN	97.95	76.69	0.45	92.73	76,69	83.95
Ensemble	99.39	98.25	0.15	97.94	93.25	95.54

Table-2.ThedetectionresultsofproposedIDS model withfourbaseclassifiersaswellaswithensemble of classifiers.

We have evaluated the performance of our proposed collaborative C-VIDS scheme with various other models like SRPDBG [22],Cross Layer [23],SPF, Watchdog, TWOACK andEAACK[24]. The following metrics were used for evaluation of the proposed C-VIDS scheme with other IDS schemes:

- Packet delivery ratio (PDR) refers to the ratio of thenumber of packets delivered to the destination node against the number of packets generated by the sourcenode.

- Routing overhead(RO*)* refers to the overhead involved in transmission due to introduction of additional routing control packets.

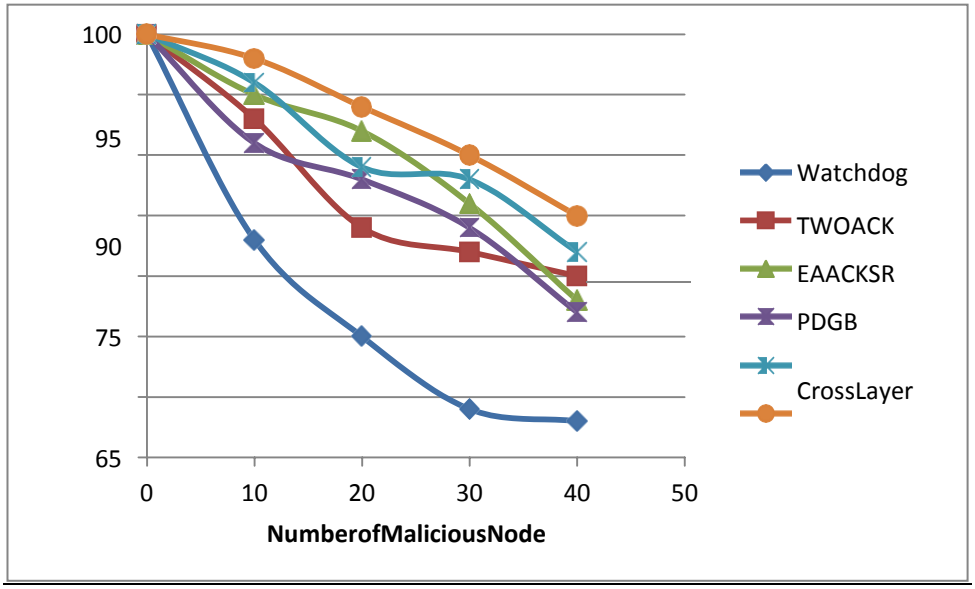

Figure-2. Packetdeliveryratio

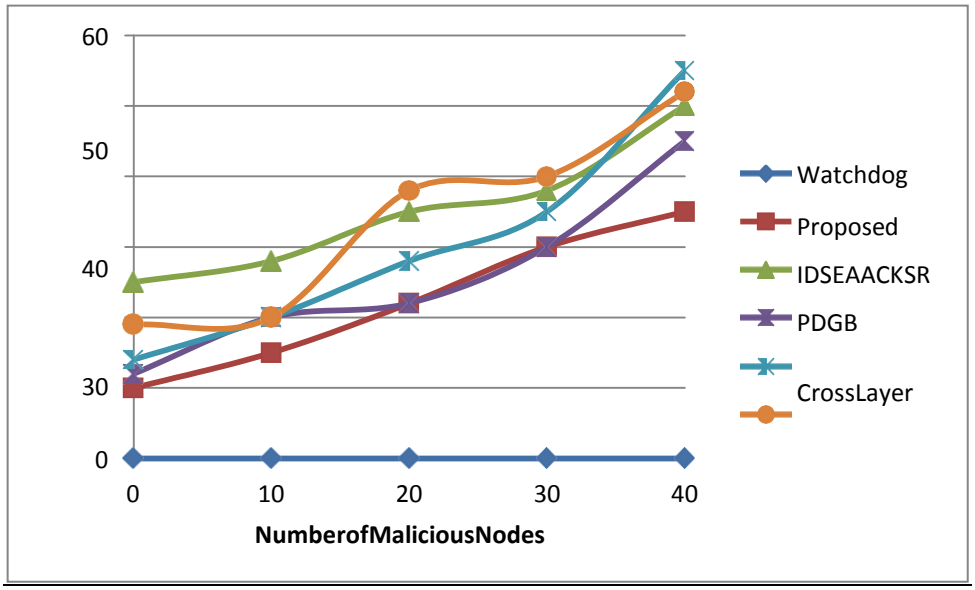

Figure-3. Routingoverhead.

Figures 2 and 3 show the *PDR* and *RO* of the various IDS schemes under varying percentage of malicious nodes. It can be observed from these figures thatall the four schemes (TWOACK, EAACK, SRPDBG and proposed IDS) have higher *PDR* than the simple WatchDog scheme. The *PDR* of our proposed IDS scheme is comparable to that of EAACK and Cross Layer schemes,while it out performs the TWOACK and SRPDBG schemes.On the otherhand, the Watchdog scheme has the least *RO*,as it does not use any acknowledgment scheme to detect misbehaving nodes.

The *RO* of the proposed IDS is less than the TWOACK, EAACK and Cross Layer schemes but higher than the SRPDBG scheme. The *RO* of the proposed IDS scheme is also better than the other existing IDS primarily due to fewer exchanges of control messages for detecting intrusion.

Conclusion and Future work:

In this work,we propose a collaborative framework for C-VIDS in VANET to detect intrusion packets and normal packets with ensemble off our classifiers.Given a packet, our collaborative C-VIDS framework will set an alarm if the packet is identified as malicious.Otherwise, it will be categorized as a normal packet. We employ ensemble off our classifiers,namely,SVM, CART,NB and K-NN with majority voting for the detection of intrusion and the normal packets.The experimental results show that our ensemble method is more robust than the other four base classifiers in the detection. In the experiments of intrusion detection, our ensemble method achieves the detection accuracy as 99.39%. The collaborative architecture is achieved with the deployment of friendly vehicle IDS nodes in the networks which are selected based on their trustworthiness.

In future work, we plan to explore the problemtrust management in a better way thereby improving theperformance of the entire intrusion detection mechanism.Designing more effective ensemble algorithms can also beinvestigated.

References:

[1]. Pathan, A.S.K. (Ed.)Security of Self-Organizing Networks:MANET, WSN, WMN, VANET; CRC Press:BocaRaton,FL,USA,2016.

[2]. Muhammad Imran, Farrukh Aslam Khan, Haider Abbas,Mohsin Iftikhar. 2019. Detection and prevention of blackhole attacks in mobile ad hoc networks, in: Proceedings ofSecurity in Ad Hoc Networks (SecAN) Workshop, 13thInternationalConferenceonAd-HocandWirelessNetworks,AdHocNow2019,and Benidorm, Spain.

[3]. Zhang, H.; Dai, S.; Li, Y.; Zhang, W. Real-time Distributed-Random-Forest-Based Network Intrusion Detection System Using Apache Spark.In Proceedings of the2 018 IEEE 37th International Performance Computing and Communications Conference(IPCCC),Orlando,FL,USA,17–19November2018;pp.1–7.

[4]. Kumar,N.;Chilamkurti,N.Collaborative trust aware intelligent intrusion detection in VANETs.Comput.Electr.Eng.**2014**,40,1981–1996.

[5]. Sharma, S.; Kaul, A. A survey on Intrusion Detection Systems and Honeypot based proactive securitymechanisms in VANETs and VANETCloud.Veh.Commun.**2018**,12, 138–164.

[6]. Liang, J.; Chen, J.; Zhu, Y.; Yu, R. A novel Intrusion Detection System for Vehicular Ad Hoc Networks(VANETs)based on differences of traffic flow and position.Appl.Soft Comput.**2019**,75,712–727.

[7]. Ghaleb, F.A.; Maarof, M.A.; Zainal, A.; Al-Rimy, B.A.S.; Saeed, F.; Al Hadhrami, T. Hybrid and MultifacetedContext-Aware Misbehavior Detection Model for Vehicular Ad Hoc Network.IEEE Access **2019**, 7,159119–159140.

[8]. Azab, A.; Layton, R.; Alazab, M.; Oliver, J. Mining malware to detect variants. In Proceedings of the 2014Fifth Cybercrime and Trustworthy Computing Conference, Auckland, New Zealand, 24–25 November 2014;pp.44–53.

[9]. Wahab,O.A.;Mourad,A.;Otrok,H.;Bentahar,J.CEAP:SVM-based intelligent detection model for clustered vehicular ad hocnetworks.ExpertSyst.Appl.**2016**, 50, 40–54.

[10]. Wu Y. S, Foo B., Mei Y., Bagchi S. 2003. Collaborative intrusion detection system(CIDS):a framework for accurate and efficient IDS. In: Proceedings of the 2003,Annual Computer Security Applications Conference(ACSAC). pp. 234-244.

[11]. Lin,X.;Sun,X.;Ho,P.-H.;Shen,X.GSIS:A Secureand Privacy-Preserving Protocol for Vehicular Communications.IEEETrans.Veh.Technol.**2007**,56,3442–3456.

[12]. Daza,V.;Domingo-Ferrer,J.;Sebé,F.;Viejo,A.Trustworthy Privacy-Preserving Car-Generated Announcements in Vehicular AdHocNetworks. IEEE Trans.Veh.Technol.**2008**,58,1876–1886.

[13]. Zhang,J.;Zulkernine,M.;Haque,A.Random-Forests-Based Network Intrusion Detection Systems. IEEE Trans Syst.Man Clyburn.PartCAppl.Rev.**2008**,38,649–659.

[14]. Huang,D.;Misra,S.;Verma,M.;Xue,G.PACP:An Efficient Pseudonymous Authentication-Based Conditional Privacy Protocol for VANETs. IEEE Trans.Intell.Transp.Syst.**2011**,12,736–746.

[15]. Shen, A.-N.; Guo, S.; Zeng, D.; Guizani, M. A lightweight privacy-preserving protocol using chamele on hashing for secure vehicular communications.In Proceedings of the 2012 IEEE Wireless Communications and Networking Conference(WCNC), Paris, France,1–4 April 2012;pp. 2543–2548.

[16]. Liu, J.K.; Yuen, T.H.; Au, M.H.; Susilo, W. Improvements on an authentication scheme for vehicular sensornetworks.ExpertSyst.Appl.**2014**, 41,2559–2564.

[17]. Li, W.;Song, H. ART: An Attack-Resistant Trust Management Scheme for Securing Vehicular Ad HocNetworks.IEEETrans.Intell.Transp.Syst.**2015**,17,960–969.

[18]. Chaubey,N.Security Analysis of Vehicular AdHoc Networks(VANETs):A Comprehensive Study.Int.J.Secur.Appl.**2016**, 10,261–274.

[19]. Yin,C.;Huang,S.;Su,P.;Gao,C.Secure routing for large-scale wireless sensor networks.In Proceedings of the International Conference on Communication Technology Proceedings,2003.ICCT2003,Beijing,China,9–11April2003;Volume2,pp.1282–1286.

[20]. Shams, E.A.; Rizaner, A.; Ulusoy, A.H. Trust aware support vector machine intrusion detection and prevention system in vehicular adhocnetworks.Comput.Secur.**2018**,78,245–254.

[21]. QuinlanJ.1986.Introduction of decisiontree,Mach.Learn.1(1): 81-106.

[22]. Kaliappan M., B. Paramasivan. 2015. Enhancing secure routing inVehicle AdHoc Networks using a Dynamic Bayesian Signalling Game model, Comput. Electr. Eng.41: 301-313.

[23]. ShresthaR, K.-H. Han D.-Y. Choi S.J. Han. 2010. A novelcross layer intrusion detection system in VANET, in: 24thIEEE International Conference on Advanced Information Networking and Applications(AINA). pp.647-654.

[24]. ShakshukiE.M,N.Kang,T.R.Sheltami.2003.EAACK-a secure intrusion-detection system for MANETs. IEEETrans.Ind. Electron. 60(3):1089-1098.