International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1266–1274

Identifying Impersonation Attack in VANET using KNN and SVM Approach

Mrugnayana S. Savekar[1], Sandeep A. Thorat[2]

[1]*Department of Computer Science & Engineering, Rajarambapu Institute of Technology, Shivaji University.*
[2]*Asst. Prof. Department of Computer Science & Engineering, Rajarambapu Institute of Technology, Shivaji University.*

Abstract

Vehicular ad hoc networks (VANETs) is useful communication in the vehicular systems using wireless technology. High mobility is one of important characteristic of VANETs. A message from one node to another node in VANET is transmitted with help of CAN (Controller Area Network) bus. VANET is prone to various security attacks. Due to lack of mechanism to verify original source and destination of the message an attacker can easily inject malicious messages in the system. This attack is called as Impersonation attack. The proposed research work uses and compares KNN and SVM machine learning algorithms to overcome Impersonation attack in VANET. The research work uses data set provided by Hacking and Countermeasure Research Lab (HCRL) for experimental evaluation. Experimental results show that KNN gives 98% accuracy for detection of Impersonation attack which is high as compared to SVM approach wherein it is 93%.

Keywords: *VANET, Intrusion Detection System, Impersonation Attack, Machine Learning, KNN, SVM.*

1. Introduction

The huge development in various modern technologies by vehicle manufacturers has improved the functions of futuristic vehicles. Now a day's cars and various types of vehicles are used by many people. A critical problem is that the transportation safety is becoming worse day by day. Even though well equipped, vehicles are dangerous. VANET [14, 15] can produce Intelligent Transportation system (ITS). The main objective of VANET is the security and flexibility. The VANET architectures are of three different types viz. WLAN (Wireless Local Area Network), pure ad-hoc, and hybrid. WLAN is connected with outside devices such as road side units so that the drivers can get real time information of vehicles. But to construct an infrastructure on road side is very time consuming and costly. Pure ad-hoc is like Vehicle to Vehicle (V2V) communication. Hybrid type is combination of WLAN and Pure Ad- hoc. An IDS (Intrusion Detection System) can identify malicious activity in vehicular communication system with the high demand to decrease the road side accidents and improving traffic security. IDS can assemble and examine the information transferred in the network to identify unauthorized access.

This research works on Classification Approach for IDS in vehicle system [3]. We used KNN and SVM algorithms to cluster intrusions. The model detects two types of attack: DoS Attack and Fuzzy Attack. The protocol used for communication between sensors and control units is named as CAN [15]. The CAN is a serial bus used for flexible performance. The connection between units is a single pair of wiring. The engine control unit, airbags, audio system etc. are the ECUs. The CAN bus protocol [20] is used to make communication easier and less complex. On board diagnostics can troubleshoot problem and create diagnosis report of the same. The speed of CAN protocol is 1 Mbps.

The proposed work identifies different attack types. This research does analysis of different parameters which gives better results than earlier methods.

Following are the key research contribution of this work:
1) The research work detects Impersonation attack using KNN and SVM.
2) The research work does experimental comparison of the different kernel functions of SVM algorithm.
3) Compare results of KNN algorithm using different values of 'K'. It also explores use of different distance functions in KNN for attack detection.

The structure of this paper is as follows: section II describes history of VANET intrusion detection system and literature survey related to proposed work. Section III describes methodology to identify attacks in the given dataset and also presents analysis of experimental results using different K-values. This section also describes comparison of different kernel functions. Section IV describes experimental environment and result using performance metrics and graphs. The paper ends by giving conclusion in the section V.

2. Literature Survey

The VANET plays the most important role in vehicular network security. We need to concentrate on the attacks possible in vehicular communication. The known attacks are identified by using the signature and encryption techniques, but they have limited prevention. To identify newer attacks we need second technique i.e. Intrusion Detection System in VANET.

Taylor *et al.* [1] describes frequency-based anomaly detection mechanism. For the classification authors used OCSVM algorithm. This can detect same information within less packet frequency insertion. Leandros A. Maglaras *et al.* [2] proposed new Distributed Intrusion Detection System (DIDS). It introduced K-OCSVM module which combine the well-known support vector machine classifier. In this module SVM with default parameters generate the outcomes which are clustered. Next, Khattab M. Ali Alheeti *et al.* [3], proposed system generates one Trace file in that all data collected and which is generated in simulation. Min-joo Kang *et al.* [4], setups DNN structure which was useful for in- vehicular network security. The model was based on Feature vector which is extract from in-vehicle network packets for which DNN can provide probability to each class to classify packets. Finally, Taylor *et al.* [5] describe anomaly-based detection mechanism which is called as LSTM recurrent neural network (RNN) to check malicious activity. The thought is to predict the future packets payload. If predicted value is different than the actual one then frame is malicious. Moayad Aloqailya *et al.* [6], in this paper author idea is to make smart vehicle communication to third parties through service request. Roland Rick *et al.* [7] describes model-based system in which sequence of events are taken into consideration for accurate prediction. The abnormal behaviour agents can be further processed using Alpha algorithm, but this model need to improve detection of attacks and payload of CAN bus events. Wang *et al.* [8] used distributed real time anomaly detection mechanism. The information is inserted into HTM module which can understand vehicular networks abnormal data sequence online. It requires co-ordination between more components. Hamid Mohammad Bhatti *et al.* [9], proposed cloud-based IDS which uses fingerprint data for the classification of activities. In the proposed techniques they assumed database which maintain the finger prints of every vehicles user. When vehicle joins RSU (Road side unit) that process verification using fingerprint then go to IDS and again it verifies behaviour of user. Next, Zhuo Weil *et al.* [10], describes vehicular IDS which is based on features of CAN message. This utilize LSTM model to train the system. Based on position of bits the classification is done. The future scope is that we need to check for another CAN IDs as well. Zhe peng *et al.* [11] describes DeepRSI. In this paper they collected data into vanet using mobile sensing technique such as GPS and data can be analyzed by smart devices such as smart phones. The proposed framework used spatio temporal relationship of vehicle GPS trajectories. Krzysztof *et al.* [12] in this paper they setup the CAN language module. Two AIDs used which detect roughly continuous messages but this model doesn't work for binary signals hence doesn't get convincing results.

There are several studies presented in literature to enhance the security of vehicular communication. This research work identifies new threat and analysis of different parameters which gives more acceptable accuracy and improves the vehicle safety.

3. Impersonation Attack Detection Using SVM and KNN

In this section, the subsection III.A introduces technical background to understand the proposed model; later subsection III.B discusses Architecture and mathematical modeling of proposed model.

A. Technical Background

Before introducing proposed mechanism, let us first briefly review the IDS with machine learning and threat types.

a) Machine Learning in IDS

IDS is more useful in network to detect malicious activities. Machine learning is mostly used in Vehicular network and for that we build IDS. K in KNN [18] is the number of instances that we used for determination. Choosing the right value of K is important for better accuracy. KNN has less calculation and accurate prediction. By using distance function, we classify new cases by using existing cases. The number of neighbors (K) [20] is a hyper parameter that you need to choose at the time of model building. In the SVM [18-19] first parameter we consider is decision boundary.

b) Threat Mechanism

Two types of attacks introduced are: DoS attack and Fuzzy attack [16]. To do analysis we used two datasets which are from real car datasets: DoS dataset and Fuzzy dataset. To insert too many requests in very short period is DoS attack. In DoS attack the '0 x 000' is the CAN ID occurs after every 0.3 millisecond [20]. In the DoS attack the machine/network shut down automatically hence that is not accessible for user. An arbitrary data is injected continuously in Fuzzy attack type. In this type injecting message occurs in every 250 milliseconds. Both datasets have the 12 attributes. We consider 1 as normal and 0 as injected.

B. Proposed Mechanism:

There are various types [14] of attacks identified in network that's why vehicle system collapses. In the existing system only, few parameters used for the analysis and predict accuracy but those are not enough to predict acceptable accuracy.

In the proposed approach, along with existing parameters we add new parameters which are kernel functions and distance functions to give better analysis results. We have identified one new attack type on same dataset.

a) Detection of Impersonation Attack

As mentioned earlier, we used raw dataset named as Impersonation Dataset which is having 12 attributes. CAN Id uniquely identifies CAN message which is generated automatically. Figure 1. Shows Detection of Impersonation attack [16]. It consists of five layers: Raw dataset, Pre-Processing, Features extraction and Normalization, classification, IDS. The following section gives detailed explanation of how raw data is processed in these layers to get appropriate output.

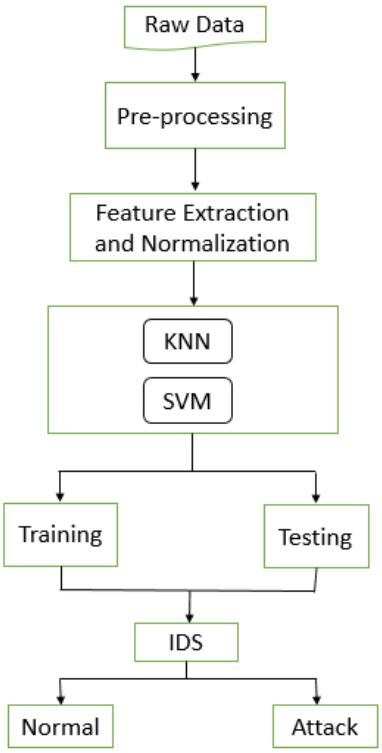

Figure 1 Detection of Impersonation Attack.

Initially we have given appropriate name to raw dataset i.e. Impersonation dataset. We pre-processed the dataset to remove unwanted columns from the dataset and rename columns i.e. add header to each column. As the CAN Id is in Hexadecimal format, we converted that in decimal format and also, we remove missing values from dataset in the Pre-Processing step.

The attributes selection process is used to select the most applicable features from the dataset and then algorithms are applied to achieve better performance of the model. χ^2-Test is used for feature extraction. We calculate χ^2 between each feature and the target and select the desired number of features with best χ^2 scores. (see Eq. 1)
χ^2 score is given by:

$$\chi^2 = \frac{(ObservedFrequeny - ExpectedFrequency)^2}{ExpectedFrequency}$$

Where,
Observed frequency = No. of observations of class,
Expected frequency = No. of expected observations of class if there was no relationship between the feature and the target.

After implementation of algorithms dataset is split in to two sections: training and testing. In KNN we calculate the distance from the item with every other neighbouring item in class. We pick the immediate K neighbours and we check for class where more items are included. We classify the new item there. We have fine-tuned the value of P and calculate the distance in two different ways. We use Manhattan Distance if we need to calculate the distance between two data points in a grid like path. In this value of P is always

set to 1. Distance d is calculated using an absolute sum of difference between its Cartesian co-ordinates as below: (see Eq. 2)

$$d = \sum_{i=1}^{n} x_i - y_i \vee$$

Where n is number of variables, x and y are the variables of vectors x and y respectively, in the two dimensional vector space. i.e. x = (x1, x2, x3,…) and y = (y1, y2, y3,...).

Euclidean distance is one of the most used distance metric. In this P value always set to 2. The distance 'd' formula as below: (see Eq.3)

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

SVM algorithms [18] use a set of mathematical functions that are defined as the kernel. To take data as input and transform it into the required form is the function of kernel. Different SVM algorithms [18] use different types of kernel functions. These functions can be of different types. For example, linear, polynomial, radial basis function (RBF), and sigmoid. Finally, the intrusion detection system generates predictions. We train the model using training dataset and predict the response for test dataset.

4. Experimental Environment and Setup

Section IV.A discusses experimental environment, such as dataset and parameters used for comparison. Section IV.B provides information about experimental results and analysis of the same. The subsection also provides information about a qualitative comparison of generated responses.

A. Dataset

Impersonation attack is evaluated using Impersonation dataset which are provided by the Hacking and Countermeasure Research Lab (HCRL) [13]. Injecting messages of Impersonating node, arbitration ID = '0x164'. It contains 695,365 numbers of messages. The dataset randomly separated into training, and testing sets with ratio of 70:30.

B. Experimental Results

In proposed mechanism we identify impersonation attack using KNN and SVM algorithms. The research work does analysis of three metrics Accuracy, Precision, and Recall [1], [20]. Other experimental analysis done by adding more parameters i.e. K-value. We test the model with K value 1 and predict with test set data and check the accuracy and other parameters then repeat the same process after increasing the k value by 1 each time. Here we have increased the k value by 1 from 1 to 10 and printing the accuracy with respected k value. Here the target is maximum accuracy, so we have taken the k value as 7 (in which accuracy was maximum after executing the code).

Table 1: Result of Nearest K-value for Impersonation Dataset

K-value	Accuracy	Precision	Recall
K=1	93	90	93
K=2	90	95	91
K=3	91	91	92
K=4	93	88	94

K=5	93	90	94
K=6	94	99	94
K=7	98	99	99
K=8	97	97	93
K=9	95	88	92
K=10	95	89	94

We have used two distance functions to check the accuracy of algorithm hence the analysis tells that if you set P=1 i.e. Manhattan Distance function which gives higher accuracy than the Euclidean Distance function.

Table 2: Comparison for Distance functions for Impersonation Dataset

Distance Function	Accuracy (in %)	Recall (in %)	Precision (in %)
Euclidean Distance	73	74	70
Manhattan Distance	74	74	72

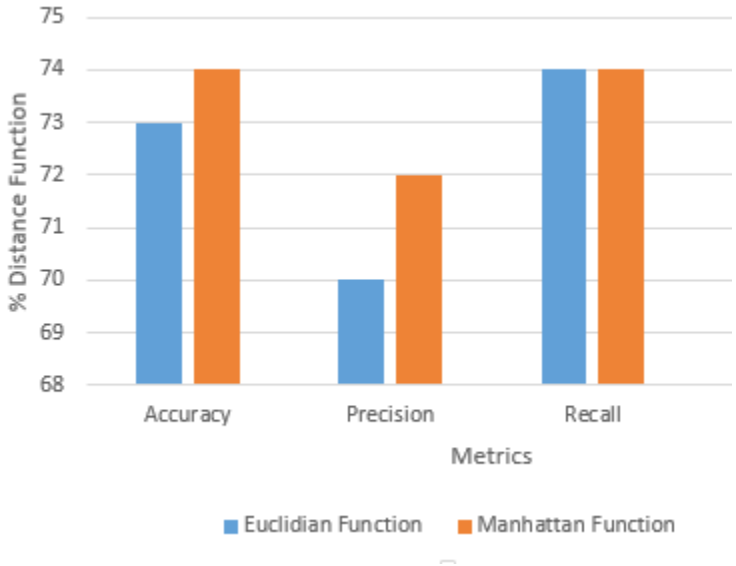

Figure 2: Comparison for Distance functions

In the SVM, Table 3 compares different kernel functions can be different types. Though analysis tells best suitable which gives more acceptable results.

Table 3: Results of metrics comparison for Kernel functions on Impersonation Dataset

Kernel Function	Accuracy (in %)	Precision (in %)	Recall (in %)
Linear	90	85	84
Poly	79	84	93
RBF	93	88	94
Sigmoid	84	88	88

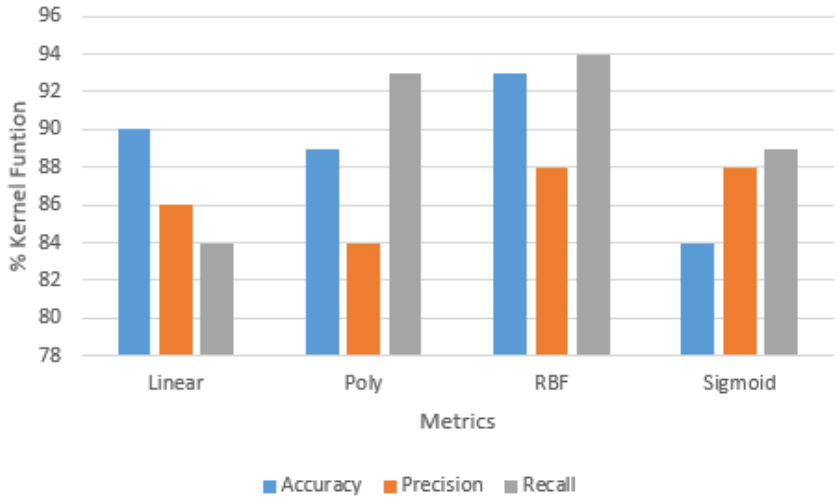

Figure 3: Results of metrics comparison for Kernel functions

The analysis report states that RBF kernel gives best accuracy than the other Kernels. Finally, research work calculate important factor which is comparative analysis of overall accuracy on Impersonation Dataset.

Table 4: Result of KNN and SVM for Impersonation attack

Algorithm	Accuracy (in %)	Precision (in %)	Recall (in %)
KNN	98	99	98
SVM	93	88	94

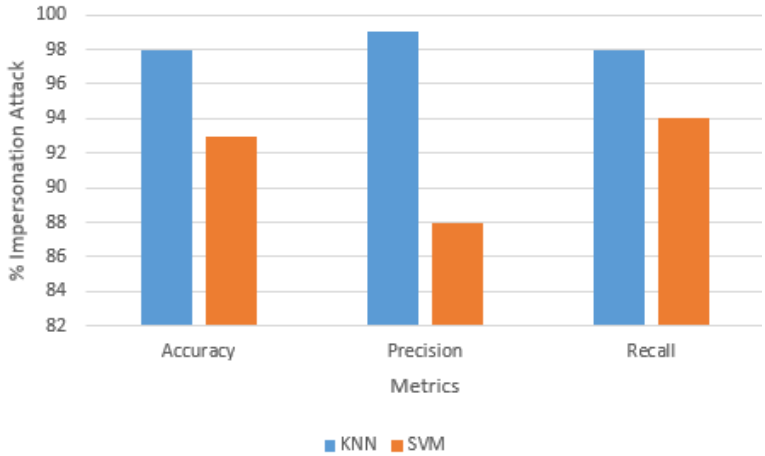

Figure 4: Comparison between SVM and KNN for detection of Impersonation attack

5. Conclusion

There are many IDS systems presented for communication between vehicles. To identify more attack types and make vehicular system more intelligent is the main task of this model. The developed model outperforms best predictions however KNN performs better which gives 98% accuracy, 99% precision and 98% recall. The research work identify that RBF kernel gives 76% accuracy which is more useful than other kernels. The Manhattan distance function gives better results than Euclidian distance function. In the future work we will implement deep learning approach and compare that with other traditional algorithms to identify more accurate results. Also we will try for multiple CAN IDs to identify more attack types.

References

[1] Taylor, Adrian, Nathalie Japkowicz, and Sylvain Leblanc. "Frequency-based anomaly detection for the automotive CAN bus." In 2015 World Congress on Industrial Control Systems Security (WCICSS), pp. 45-49. IEEE, 2015.

[2] Riebl, Raphael, Markus Monz, Simon Varga, Helge Janicke, Leandros Maglaras, Ali Hilal Al-Bayatti, and Christian Facchi. "Improved security performance for vanet simulations." 4th IFAC Symposium on Telematics Applications, 2016.

[3] Alheeti, Khattab M. Ali, Anna Gruebler, and Klaus D. McDonald-Maier. "An intrusion detection system against black hole attacks on the communication network of self-driving cars." In 2015 sixth international conference on emerging security technologies (EST), pp. 86-91. IEEE, 2015.

[4] Kang, Min-Joo, and Je-Won Kang. "Intrusion detection system using deep neural network for in-vehicle network security." PloS one 11, no. 6 (2016).

[5] Taylor, Adrian, Sylvain Leblanc, and Nathalie Japkowicz. "Anomaly detection in automobile control network data with long short-term memory networks." In 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 130-139. IEEE, 2016.

[6] Aloqaily, Moayad, Safa Otoum, Ismaeel Al Ridhawi, and Yaser Jararweh. "An intrusion detection system for connected vehicles in smart cities." Ad Hoc Networks 90 (2019): 101842.

[7] Rieke, Roland, Marc Seidemann, Elise Kengni Talla, Daniel Zelle, and Bernhard Seeger. "Behavior analysis for safety and security in automotive systems." In 2017 25th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), pp. 381-385. IEEE, 2017.

[8] Wang, Chundong, Zhentang Zhao, Liangyi Gong, Likun Zhu, Zheli Liu, and Xiaochun Cheng. "A distributed anomaly detection system for in-vehicle network using HTM." IEEE Access 6 (2018): 9091-9098.

[9] Lokman, Siti-Farhana, Abu Talib Othman, and Muhammad-Husaini Abu-Bakar. "Intrusion detection system for automotive Controller Area Network (CAN) bus system: a review." EURASIP Journal on Wireless Communications and Networking 2019, no. 1 (2019): 184.

[10] Wei, Zhuo, Yanjiang Yang, Yasmin Rehana, Yongdong Wu, Jian Weng, and Robert H. Deng. "IoVShield: an efficient vehicular intrusion detection system for self-driving (short paper)." In International Conference on Information Security Practice and Experience, pp. 638-647. Springer, Cham, 2017.

[11] Sheikh, Muhammad Sameer, Jun Liang, and Wensong Wang. "A Survey of Security Services, Attacks, and Applications for Vehicular Ad Hoc Networks (VANETs)." Sensors 19, no. 16 (2019): 3589.

[12] Wang, Chundong, Zhentang Zhao, Liangyi Gong, Likun Zhu, Zheli Liu, and Xiaochun Cheng. "A distributed anomaly detection system for in-vehicle network using HTM." IEEE Access 6 (2018): 9091-9098.

[13] Hacking and Countermeasure Research Lab (2017) CAN-Intrusion-Dataset. http://ocslab.hksecurity.net/Dataset/CAN-intrusion-dataset
[14] HAMADA, Yoshihiro, Masayuki INOUE, Naoki ADACHI, Hiroshi UEDA, Yukihiro MIYASHITA, and Yoichi HATA. "Intrusion Detection System for In-Vehicle Networks." SEI TECHNICAL REVIEW 88 (2019): 77.
[15] Dupont, Guillaume, Jerry den Hartog, Sandro Etalle, and Alexios Lekidis. "Network intrusion detection systems for in-vehicle network-Technical report." arXiv preprint arXiv:1905.11587 (2019).
[16] La, Vinh Hoa, and Ana Rosa Cavalli. "Security attacks and solutions in vehicular ad hoc networks: a survey." (2014).
[17] Peterson, Leif E. "K-nearest neighbor." Scholarpedia 4, no. 2 (2009): 1883.
[18] Joachims, Thorsten. Learning to classify text using support vector machines. Vol. 668. Springer Science \& Business Media, 2002.
[19] Bhatia, Nitin. "Survey of nearest neighbor techniques." arXiv preprint arXiv:1007.0085 (2010).
[20] Alshammari, Abdulaziz, Mohamed A. Zohdy, Debatosh Debnath, and George Corser. "Classification Approach for Intrusion Detection in Vehicle Systems." Wireless Engineering and Technology 9, no. 4 (2018): 79-94.