A lightweight Anomaly based DDoS flood attack detection for Internet of vehicles

Kuthada Mohan Sai*[1], Brij .B Gupta[2], Francesco COLACE[3], Kwok Tai Chui*[4]

[1]*Department of Computer Engineering, National Institute of Technology Kurukshetra, India*
[2]*National Institute of Technology Kurukshetra, Kurukshetra, Haryana 136119, India, & Asia University, Taichung 413, Taiwan & Staffordshire University, Stoke-on-Trent ST4 2DE, UK*
[3] *University of Salerno, Italy*
[4] *Hong Kong Metropolitan University (HKMU) , Hong Kong*

*Corresponding Author

Abstract

The concept of the Internet of Vehicles (IoV) enhances the VANETs by merging the VANETs with the Internet of things (IoT) thus making intelligent transportation systems a reality. The intelligent transport systems generate greater volumes of critical dynamic real-time data and thus raise a concern in the security of the generated data. The IoV has become a prominent field because of its scalability, reliable internet connection, and dynamic topological structures and due to its compatibility with various devices and sensors. IoV is susceptible to a range of attacks. The IoV consists of various kinds of components which involve various communications with sensors, vehicles, road infrastructure and humans. This paper will focus on UDP based Distributed Denial of Service (DDOS) Flood attacks. Onboard unit (OBU) is a computational device present in the vehicle is a resource-constrained device a lightweight DDoS detection machine learning algorithm is required to detect the DDoS attack performed on the vehicles by a dataset generated using OMNET++ simulator.

Keywords

IoV, DDoS, IoT, Machine learning, SVM, J48

1. Introduction

With the exponential increase of usage of vehicles and changing relationships among the vehicles, their communications raise a concern in the security of their communication as they are prone to become highly complex. IoT is described as connected smart devices that interact with the environment surrounding them. Many advances of the IoT, are integrated into vehicles to model Intelligent Transportation System (ITS). The IoV is a subset of IoT in which the data or the information generated is shared among vehicles, pedestrians, infrastructures, and traffic lights. A reliable and efficient communication service can be achieved by implementing the IoV [1]. In order to realize the concepts of smart cities ITS techniques have to be employed. The 5G enabled communications has been employed to IoV to increase the bandwidth for the users thus resulting in more amount of information being transmitted in the IoVs. Moreover, the network of the IoV is an open network in which any user who might be an attacker can also join the network. The networks of IoV are open because that the IoV networks are prone to attacks that are caused by the worm and the Trojan horses and cyber-attacks [1].

There are 5 types of heterogeneous communication involved when the vehicle is communicating with the surrounding objects they are of Vehicle-to-Vehicle (V2V), Vehicle-to-Roadside Units (V2R), Vehicle-to-Personal devices (V2P), Vehicle-to-Sensors (V2S), and Vehicle-to-Infrastructure of networks (V2I) Fig. 1. The data or the information generated by the IoV is categorized into two: on-

International Conference on Smart Systems and Advanced Computing (Syscom-2021), December25–26, 2021
EMAIL: mohankuthada@gmail.com, gupta.brij@gmail.com, fcolace@unisa.it, jktchui@ouhk.edu.hk
© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3080/20.pdf

road and onboard data. Vehicle's status like brakes, the velocity of the vehicle and working of various components of the engine are monitored by onboard data. The functionalities that are carried out on the road such as distance between the vehicles, pilot camera, blind spots and the traffic lights [4] are monitored by the on-road data. To perform the computations an onboard unit (OBU) is present within the vehicle. It is also used to carry out Inter as well as intra communications that are depicted in Figure 1.

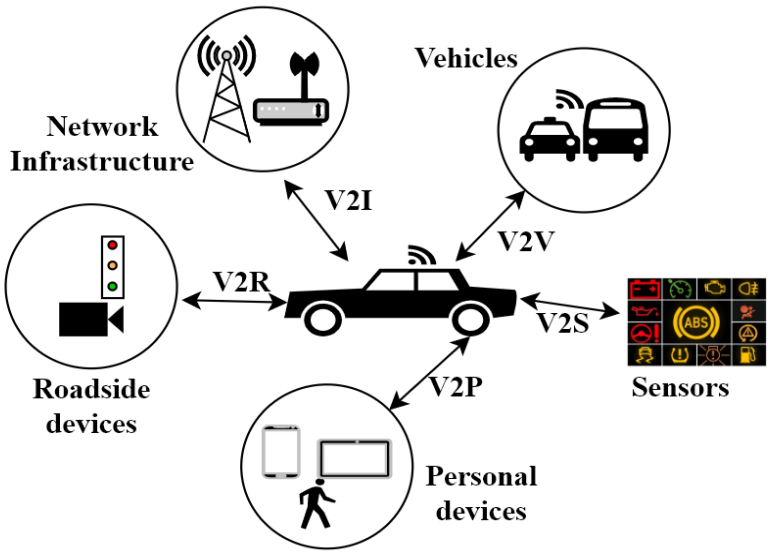

Figure 1: Heterogeneous Communication in IoV

The Distributed Denial of Service (DDoS) [2] attack is one of the major threats for vehicles present in the IoV. Among the taxonomy of DDoS attacks flooding attacks are of major concern as they exhaust the cache and the computing resources of the OBU present on the vehicles. DDoS attack creates huge traffic using the DDoS attack from which the attacker exhausts the resources of the targeted host like the network bandwidth and the CPU time[26].

Some of the examples of DDoS attacks are DNS flood, SYN Flood, UDP flood and Ping Flood [5][6]. The OBU of the vehicles is connected to various devices through wired and wireless means where each and every component is connected globally.

As the devices are connected globally and the resources of these devices have highly controlled the security of IoV becomes highly challenging[3]. In our paper, we implement an anomaly-based lightweight DDoS detection, we generated a dataset using the OMNET++[19], sumo[20] veins[22] and INET[21] tools in order to generate a dataset for the UDP flood attack and for reducing the number of features and an efficient feature selection algorithm Correlation-based feature selection algorithm is used to train a machine-learning algorithm Support vector machine(SVM) and the J48 classifier to classify the incoming traffic as positive or negative. Many researchers have proved that the SVM classifier has outperformed other classifier algorithms like k- nearest neighbour, random forest and even neural networks [8] in order to verify this claim we used the J48 classifier to compare it with SVM.

2. Related work

Recently, authors propsed different techniques for DDoS attack detection [7][13][10]. Authors of [9] proposed an anomaly detection scheme for VANETs in which they have used a convolution neural network for extracting the feature of the network traffic and to distinguish the incoming traffic they have used a threshold-based separation method.

This approach [11] is an anomaly-based DDoS detection approach that uses Cauchy distribution based optimization for the feature selection and uses SVM based classifier for the attack classification.

The authors of [12] presented an intelligent transport system using visual analytics in order to detect intrusions in road traffic data by integrating various data mining algorithms.

The authors of this model [25] used the Correlation-based feature selection for reducing the number of features to 7 and performed the classification using a machine learning algorithm called J48 classifier on a Raspberry Pi and achieved satisfactory results for an IoT scenario.

The support vector machine-based IDS proposed in [27] uses 3 features that are extracted from the packet arrival rate. Performance evaluation of this approach is done on a MatLab simulation and achieved satisfactory results for the DoS attack.

In [28], in order to find out the anomalous nodes in the IoV network, the authors have proposed a statistical approach IDS for analysing the traffic flows. The model analyses the traffic and based on the insights provided by the model it classifies as an attack or not. The model can detect the anomalous nodes and attacks in the network.

3. Motivation

The works proposed in the above approaches were designed using the datasets which were not suitable for the IoV and VANET scenarios. So a dataset especially created for the DDoS for the IoV scenario is much needed. Also, the works proposed earlier were focused on VANETs and hardly there is research done in the field of IoV for detection of cyberattacks in IoV. These things were considered in order to move ahead with our approach.

4. Proposed approach and Implementation

A dataset is needed in order to design an anomaly-based approach for the IoV and the dataset should be related to the IoV scenarios. To get UDP based DDoS dataset we used simulators to generate the data. This dataset will be used to train the machine learning model which is deployed in the OBU of the vehicles. The tools which were used and their functionality are given in Table 1. A realistic test bed is required in order to get a perfect dataset. The SUMO tool is used for generating mobility traffic which is realistic. This traffic is imported into the OMNET++. The OMNET++ will use the INET libraries and VEINS framework to generate the network traffic in which the normal and attack packets for the IoV scenarios are configured. The resulting simulation generates scalar and vector files that keep track of every instance that took place during the simulation. These result files are filtered and labelled using the various SQL and normalizing techniques to obtain the required dataset.

Table 1
Tools used

Features	Description
Throughput	Amount of data for a given amount of time
RepackHigh	Number of packets received from Higher layers
RepackLow	Number of packets received from Lower layers
PacketC	Count of the Packets

The implementation flow of the proposed approach is given in Figure 2.

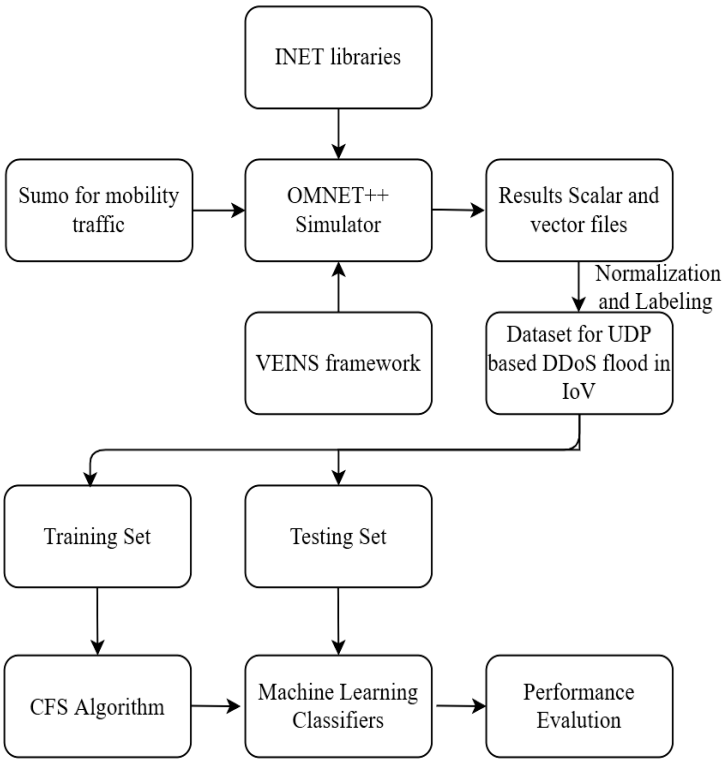

Figure 2: Overview of implementation

The resulting dataset has 29 features with 90350 instances which make it a heavy anomaly model for the OBU to process as it is power and resource-constrained. Hence it is necessary to make the model lightweight. We used a correlation-based feature selection model [15] to select the features that will be sufficient for training the model. In order to evaluate the performance of the dataset, we used J48 [18] and SVM[14] classifier machine learning algorithms. In order to evaluate the performance of the machine learning models, we used Raspberry Pi3b+ [23] which is a resource and power-constrained device like the OBU present in the vehicles. Brief information about the CFS and J48, SVM classifiers is given below.

4.1. Correlation based feature selection

"The algorithm evaluates the relation between output and correlated inputs"[15]. The features for a classifier machine learning model can be selected based on the correlation among the features. The irrelevant and redundant features are ignored by this algorithm. The redundancy of a feature is determined by the algorithm when it is correlated with one feature or more features. The subset of features is selected when they are highly related to class and not correlated with each other. To improve the accuracy and for reducing the computation of the machine learning model the features that are redundant and not relevant are ignored. A subset of features with the highest merit is selected as the feature set and the selected features are utilised in order to train the proposed model. The merit of a feature subset is given by equation 1. In the equation the number of features present in a given subset is defined by k, the average of feature-class correlation is given by r_{cf} and the average of feature-feature correlation is given by r_{ff}.

$$Ms_k = \frac{k\bar{r}_{cf}}{\sqrt{k+k(k-1)\bar{r}_{ff}}} \quad (1)$$

4.2. J48 Classifier

It is a machine learning classifier algorithm proposed by Ross quinaln[18] and it is one of the decision tree generation algorithms. The J48 classifier is also a statistical classifier. The structure of the J48 classifier is a tree-like structure that has root nodes and leaf noodles which consists of the attributes or classes. The main advantage of J48 is that it can handle both continuous and discrete features. The overfitting problem in J48 can be solved by using the pruning techniques. The J48 classifier can also be used for datasets that do not have complete data in their datasets or even with non-significant features.

4.3. Support vector machine(SVM)

"The support vector machine (SVM) is a supervised machine learning algorithm which is used for both classification and regression purposes"[14]. Wide ranges of applications use SVM as a classifier. The working of SVM is based on the support vectors and hyperplane.

A hyperplane is a flat subspace that has one less dimension of the coordinate system that it is represented. For a 2-d coordinate system it is given by $P_0 + P_1X_1 + P_2X_2 = 0$ In m-D coordinate it is represented by $P_0 + P_1X_1 + P_2X_2 + \ldots\ldots + P_mX_m = 0$.

The CFS algorithm is applied to the dataset generated from the simulation with the help of the weka [24] evaluation tool and the top 4 features from the resulting feature set were chosen to train the machine learning algorithms. The machine learning models were trained and tested in the weka evaluation tool using the generated dataset that is divided into two parts one consisting of 63245 training instances and the other consisting of 27105 testing instances. By using the CFS algorithm the features are reduced to 4 and all the instances of the training data set were handled by the Raspberry Pi and the dataset was used for evaluating the model on the WEKA tool by using the SVM and J48 classifiers.

The performance parameters of the model with SVM and J48 are compared and observed which is given in section 5. The parameters used for evaluating are the confusion matrix, F-measure, recall, precision, "False positive rate (FPR), True positive rate (TPR), False Negative rate (TNR), True Negative rate (TNR) and accuracy".

5. Results and observations

5.1. Results

The dataset generated by the OMNET++ simulator was fed into the CFS algorithms and the top four features of the selected feature set by the CFS algorithm are given in the Table 2 along with their description.performance parameters of the model with SVM and J48 are compared and observed which is given in this section.

Table 2
Features selected by CFS

Features	Description
Throughput	Amount of data for a given amount of time
RepackHigh	Number of packets received from Higher layers
RepackLow	Number of packets received from Lower layers
PacketC	Count of the Packets

The confusion matrixes for the J48 and SVM classifier are given in Figure 3 and 4 respectively.

A classified instance is said to be true positive if the classified instance is an attack packet. An instance is said to be true negative if the classified instance is a normal packet. An instance is said to be false positive if the instance is a normal packet but classified as attack packet. An instance is said to be false negative if the instance is an attack packet but classified as normal packet.

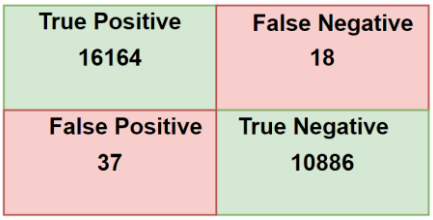

Figure 3: Confusion matrix for J48

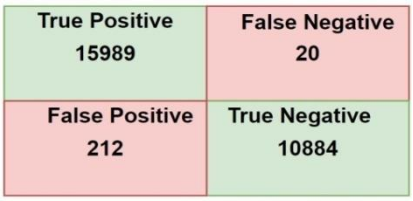

Figure 4: Confusion matrix for SVM

The detection accuracy of the trained models is compared in Table 3 by evaluating the models in WEKA evaluation tool running on Raspberry Pi 3b+.

Table 3
Detection accuracy of proposed model with J48 and SVM

Measure	Value for J48	Value for SVM
Sensitivity	0.9989	0.9988
Specificity	0.9966	0.9809
Precision	0.9977	0.9869
Negative Predictive Value	0.9983	0.9982
False Positive Rate	0.0034	0.0191
False Negative Rate	0.0011	0.0012
Accuracy	0.9980	0.9914
F1 score	0.9983	0.9928

5.2. Observations

- By using CFS we have achieved lightweight DDoS detection for OBU by greatly reducing the number of features. It is known that if the number of features are more, the system becomes more complex and it requires more computational power. By reducing the number of features from 29

to 4 we can observe that the complexity is greatly reduced and hence our system uses less power and resources for computation.

- It can also be seen that the J48 classifier provided better detection accuracy when compared to SVM as the false positive rate of the J48 is much lower than the false positive rate of SVM. Hence we can say that J48 classifier can be used as the machine learning model to design the anomaly-model for the IoV.

6. Conclusion and future plan

The IoV is an emerging technology and becoming a reality and due to its open network nature, it is vulnerable to cyber-attacks. In this paper, we implemented a lightweight anomaly-based DDoS detector to distinguish the network traffic from various devices to OBU in order to safeguard the OBUs. To achieve this in a lightweight scenario CFS algorithm is used for reducing the feature from the dataset we have generated using the OMNET++ simulator which has the instances for normal and UDP based DDoS flood attacks. The features are reduced from 29 to 4; these four features are used to train the machine learning Algorithms J48 and SVM on the test dataset. The models were evaluated on the WEKA evaluation tool which is installed on Raspberry pi 3b+. The models were compared according to their detection accuracy and it is evident that the J48 classifier outperformed the SVM classifier and the CFS algorithm has made the system lightweight by reducing the number of features. The applications for the proposed model can be further extended into various scenarios of the Internet of Everything (IoE) such as healthcare, smart cities, Agriculture, Industries, and irrigation. For future work, we can extend the instances in the dataset for various cyber-attacks which can be used to model an intrusion detection system for the Internet of vehicles.

References

[1] "B. B. Nie, L., Ning, Z., Wang, X., Hu, X., Li, Y., & Cheng, J. (2020). Data-Driven Intrusion Detection for Intelligent Internet of Vehicles: A Deep Convolutional Neural Network-based Method. IEEE Transactions on Network Science and Engineering.

[2] Srivastava, A., Gupta, B. B., Tyagi, A., Sharma, A., & Mishra, A. (2011, September). A recent survey on DDoS attacks and defense mechanisms. In International Conference on Parallel Distributed Computing Technologies and Applications (pp. 570-580). Springer, Berlin, Heidelberg.

[3] A. Dahiya, B. B. Gupta (2021) How IoT is Making DDoS Attacks More Dangerous?, Insights2Techinfo, pp.1

[4] Sherazi, H. H. R., Iqbal, R., Ahmad, F., Khan, Z. A., & Chaudary, M. H. (2019). DDoS attack detection: A key enabler for sustainable communication in internet of vehicles. Sustainable Computing: Informatics and Systems, 23, 13-20

[5] Dhananjay Singh (2021) Captcha Improvement: Security from DDoS Attack, Insights2Techinfo, pp.1

[6] Salim, M. M., Rathore, S., & Park, J. H. (2019). Distributed denial of service attacks and its defenses in IoT: a survey. The Journal of Supercomputing, 1-44.

[7] Gupta, B. B., Joshi, R. C., & Misra, M. (2012). ANN Based Scheme to Predict Number of Zombies in a DDoS Attack. Int. J. Netw. Secur., 14(2), 61-70.

[8] Jan, S. U., Ahmed, S., Shakhov, V., & Koo, I. (2019). Toward a lightweight intrusion detection system for the internet of things. IEEE Access, 7, 42450-42471.

[9] Laisen Nie, Yongkang Li, and Xiangjie Kong. 2018. Spatio-Temporal Network Traffic estimation and Anomaly Detection Based on Convolutional Neural Network in Vehicular Ad-Hoc Networks. IEEE Access 6 (2018), 40168–40176.

[10] Chhabra, M., Gupta, B., & Almomani, A. (2013). A novel solution to handle DDOS attack in MANET. Journal of Information Security Vol. 4 No. 3 (2013) , Article ID: 34631 , 15 pages DOI:10.4236/jis.2013.43019

[11] Garg, S., Kaur, K., Kaddoum, G., Gagnon, F., Kumar, N., & Han, Z. (2019, July). Sec-IoV: A multi-stage anomaly detection scheme for Internet of vehicles. In Proceedings of the ACM Mobi-Hoc Workshop on Pervasive Systems in the IoT Era (pp. 37-42).
[12] Maria Riveiro, Mikael Lebram, and Marcus Elmer. 2017. Anomaly Detection for Road Traffic: A Visual Analytics Framework. IEEE Transactions on Intelligent Transportation Systems 18, 8 (2017), 2260–2270.
[13] Gupta, B. B., Misra, M., & Joshi, R. C. (2012). An ISP level solution to combat DDoS attacks using combined statistical based approach. arXiv preprint arXiv:1203.2400.
[14] Rossi, F., Villa, N. (2006). Support vector machine for functional data classification. Neurocomputing, 69(7-9), 730-742.
[15] Hall, M. A. (1999). Correlation-based feature selection for machine learning.
[16] Search UNB. (n.d.). Retrieved November 18, 2020, from https://www.unb.ca/cic/research/applications.html
[17] Sharafaldin, I., Lashkari, A. H., Hakak, S., & Ghorbani, A. A. (2019, October). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In 2019 International Carnahan Conference on Security Technology (ICCST) (pp. 1-8). IEEE.
[18] Ashari, A., Paryudi, I., & Tjoa, A. M. (2013). Performance comparison between Naïve Bayes, decision tree and k-nearest neighbor in searching alternative design in an energy simulation tool. *International Journal of Advanced Computer Science and Applications (IJACSA), 4*(11).
[19] *OMNeT++ Discrete Event Simulator*. (n.d.). OMNET++. Retrieved May 10, 2021, from https://omnetpp.org/
[20] Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y. P., Hilbrich, R., ... & Wießner, E. (2018, November). Microscopic traffic simulation using sumo. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)* (pp. 2575-2582). IEEE.
[21] *INET Framework - INET Framework*. (n.d.). INET. Retrieved May 10, 2021, from https://inet.omnetpp.org/
[22] *Documentation - Veins*. (n.d.-b). VEINS. Retrieved May 10, 2021, from https://veins.car2x.org/documentation/
[23] Raspberry Pi 3b+ datasheet. Retrieved May 10, 2021 from https://static.raspberrypi.org/files/product-briefs/Raspberry-Pi-Model-Bplus-Product-Brief.pdf
[24] Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (n.d.). WEKA. Retrieved May 10, 2021, from *https://www.cs.waikato.ac.nz/ml/weka/*
[25] Soe, Y. N., Feng, Y., Santosa, P. I., Hartanto, R., & Sakurai, K. (2019, March). Implementing Lightweight IoT-IDS on Raspberry Pi Using Correlation-Based Feature Selection and Its Performance Evaluation. In *International Conference on Advanced Information Networking and Applications* (pp. 458-469). Springer, Cham
[26] Tripathi, S., Gupta, B., Almomani, A., Mishra, A., & Veluru, S. (2013). Hadoop based defense solution to handle distributed denial of service (ddos) attacks. Journal of Information Security. Vol. 4 No. 3 (2013) , Article ID: 34629 , 15 pages DOI:10.4236/jis.2013.43018
[27] Jan, S. U., Ahmed, S., Shakhov, V., & Koo, I. (2019). Toward a lightweight intrusion detection system for the internet of things. *IEEE Access*, *7*, 42450-42471.
[28] Zaidi, K., Milojevic, M. B., Rakocevic, V., Nallanathan, A., & Rajarajan, M. (2015).Host-based intrusion detection for vanets: a statistical approach to rogue node detection.IEEE transactions on vehicular technology, 65(8), 6703-6714.