**FEATURE SELECTION AND CLASSIFICATION METHOD FOR ATTACK DETECTION USING DEEP LEARNING IN VANET**

**NAME: P. SRILAKSHMI**

**REG NO: 2019272037**

**CLASS: MCA-SS**

**GUIDE: Ms. G. MAHALAKSHMI**

**ABSTRACT:**

Vehicular Ad hoc Networks (VANETs) are the core of Intelligent Transportation Systems (ITS), allowing vehicles to communicate between themselves and with other entities. However, these are very complex networks with volatile architectures, ever-changing members, and multiple types of entities, making them an appealing target for attackers since they can find vulnerabilities and perform attacks with massive impact.

The presence of a large number of communicating vehicles greatly increases the number and types of possible anomalies in the network. However, since there can be a multitude of anomalies possible in the network, there is a need for better anomaly detection frameworks that can address this unprecedented scenario. In this project, we propose an anomaly detection framework for VANETs based on deep neural networks (DNNs) using PCA and DNN algorithms. Our DNN model learns the abstract and high-dimensional feature representation of the IDS data by passing them into many hidden layers.

**INTRODUCTION:**

A **Vehicular ad hoc network** called **VANET** is a mobile network allowing to vehicles to communicate with each other in the absence of fixed infrastructure, with the aim of improving road safety through the exchange of alerts between vehicles. A VANET can be utilised to provide Peer to peer application, Internet connectivity and other services for the user apart from safety. Peer to peer applications is useful to provide services like sharing music, movies etc. among the vehicles in the network. People always want to connect with the Internet all the time, hence VANET provides the constant connectivity of the Internet to the users. The security in VANET is most critical issue because the information is propagated in open access environment. VANET's are exposed to various threats and attacks. It is necessary that all the data which is transmitted should not be changed by the attackers. So, the ultimate goal of all works toward VANET is to provide road safety information among the nodes hence the frequent exchange of such type of data on the network clearly signifies the role of the security. An effective way to identify when an attack occurs in a VANET is the deployment of an Intrusion Detection System (IDS).

An **intrusion detection system (IDS**) is a mechanism to identify abnormal or suspicious activities on the target network. In this project, we identify the attacks and classify them accordingly using deep learning algorithms. There are many types of attacks in VANET such as Denial of Service (DoS) attack, distributed Denial of Service (DDoS) attack, Sybil attack, Application attack, Timing attack, etc. Deep learning algorithms have significant attention and  
been widely used in various fields to improve the performance of the previous methods. Deep learning-based methods can automatically extract and selects features using raw data. For proposed intrusion detection system in VANET to detect intrusion deep belief network algorithm of deep learning will use. It is an effective method of solving the problems from neural network with deep layers.

**Deep Learning (DL)** has gained significant attention and been widely used in various fields to improve the performance of the previous methods. DL-Based methods can obviate the burdensome work of selecting features and acquiring private information since it automatically extracts and selects features using raw data. On the perspective of resource requirement, although training a DL-Based method requires a high computational resource, yet most of the trained DL classifier is of small size and computational economic. In a nutshell, DL-Based methods can attain a higher performance.

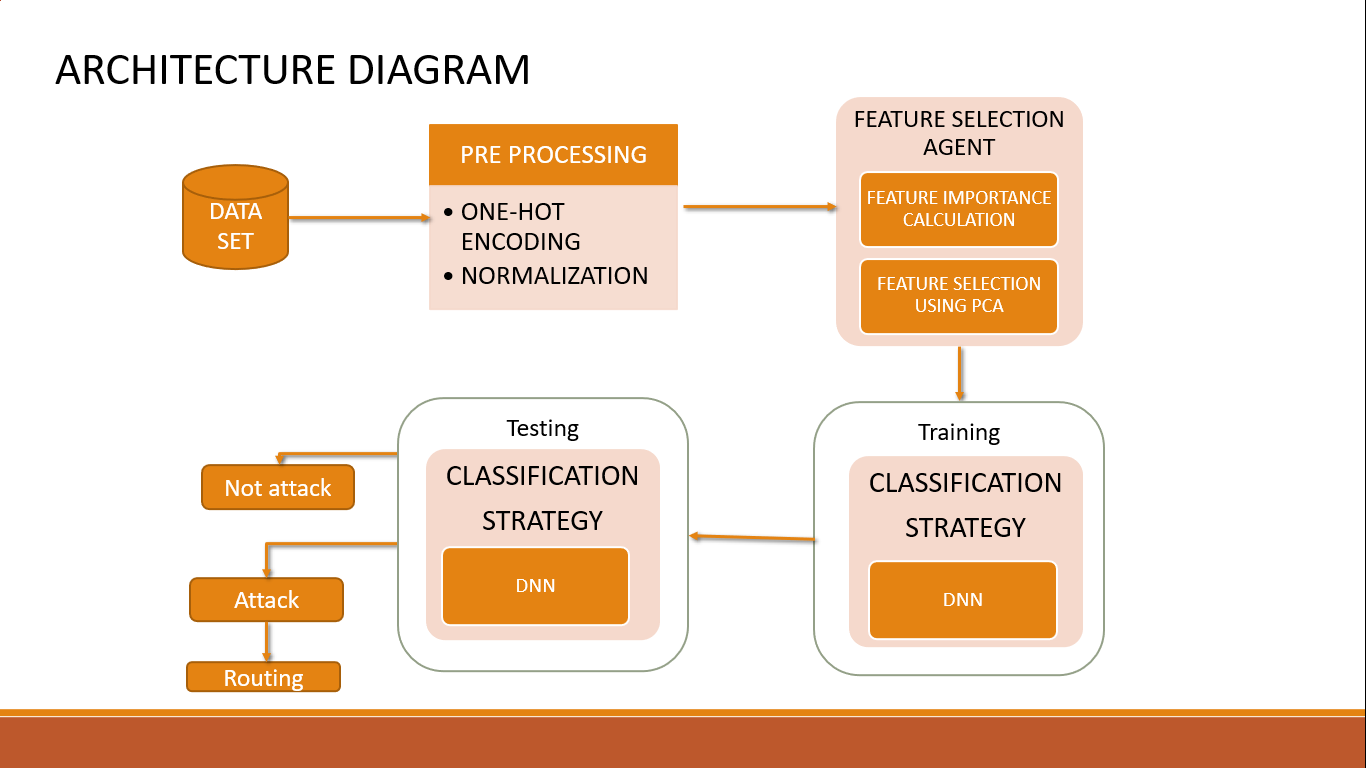
**PROBLEM STATEMENT:**

This project aims to detect and identify the different types of attacks that occur in the VANETs. We use Deep learning-based intrusion detection system. In this project we present an end-to-end DL-Based intrusion detection method for the VANET with only the given datasets. We also evaluate our method on two public datasets, experimental results show the capability of applying this in the VANET and the effectiveness of detecting malware traffic.This project uses DNN algorithm. They are structured in the form of interconnected nodes with an input layer, an output layer and one or more hidden layers. The model as various layers where input to each node is produced by applying some non-linear function on the data.

**OBJECTIVE:**

The aim of our study is to incorporate feature engineering to enhance the performance of Deep Learning techniques for attack classification. Due to the wide variety of attacks, and the unavoidable loopholes in traditional security methods, it is not possible to entirely prevent attacks on VANETs. This is done using the available feature selection and classification algorithm through which we get best results for the dataset. Numerical results will show that we can correctly detect data corresponding to several anomaly types. In this project we concentrate on timing attacks.

**ARCHITECTURE DIAGRAM:**



**ARCHITECTURE EXPLANATION:**

**Load Dataset:**

We gather the datasets KDDCup99 and USNW-NB15. The USNW-NB15 dataset has nine types of attacks, namely, Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode and Worms. This has 49 features with the class label. A partition from this data set is configured as a training set and testing set, namely, UNSW NB15 training-set.csv and UNSW\_NB15\_testing set respectively. The number of records in the training set is 175,341 records and the testing set is 82,332 records from different the types of attack and normal. In KDDcup99, some intrusion experts believe that most novel attacks are variants of known attacks and the "signature" of known attacks can be sufficient to catch novel variants. The datasets contain a total of 24 training attack types, with an additional 14 types in the test data only.

**Pre-Processing:**

This step involves the cleaning of the data set by removing the null values, noisy data and duplicate values. This is an important step as it ensures low repetition and consistency in data. In this project, encoding and normalization is done to pre-process the dataset.

**Feature Selection Agent:**

Feature selection agent is the process of reducing the dimension of our dataset by selecting only the necessary features. This is done using the feature importance calculations which calculates the information gain and entropy value. The we perform the feature selection using the PCA (Principal Component Analysis) method.

**Training and Testing:**

Training and Testing is necessary for us to know if the said model works. The observations in the training set form the experience that the algorithm uses to learn. The test set is a set of observations used to evaluate the performance of the model using some performance metric. It is important that no observations from the training set are included in the test set. In this DNN classification algorithm is used to train and test the dataset and check if the output we obtain is accurate. The accuracy can be tested using the accuracy formula.

**MODULE AND MODULE DESCRIPTION:**

**Modules:**

* Pre-processing the data.
* Feature Selection.
* Training and testing the classification algorithm.

**Brief module description:**

Pre-processing data:

The first sept is to pre-process the dataset using encoding and normalization techniques. The dataset is composed of various types of fault and attack type messages which are together considered as Anomaly class in this project. It is aimed to have low repetition and a large sample size for our customized dataset. After this, the missing data and the meaningless data that don’t pertain to the dataset are removed.

Then it is encoded data using one hot encoding technique. One-Hot-Encoding is used to convert all categorical properties to binary properties. One-Hot-Encoding requirement, the input to this transformer must be an integer matrix expressing values taken with categorical (discrete) properties. The normalization of the dataset is done accordingly. The data normalization is a basic element of data mining. It means transforming the data, namely converting the source data in to another format that allows processing data effectively. It uses the min-max techniques to normalize the dataset. The main purpose of data normalization is to minimize or even exclude duplicated data.

Feature Selection:

To secure the network against intrusions, real-time intrusion detection is required that can be achieved by finding significant features from available dataset. The reduced set of features can efficiently increase the detection rate of intrusions. This is done using feature selection techniques. Selecting the features is the process of identifying subset of features from the available dataset. This is achieved by evaluating the features on the basis of some criteria such as information gain and entropy values.

Feature selection minimizes the size of the dataset that is further given as an input to the classification algorithm. Selecting features enhances the understand-ability and visualization of learning algorithm for intrusion detection. In this project, Principal Component Analysis (PCA) is used for feature selection. Principal Component Analysis, or PCA, is a dimensionality-reduction method that is often used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information in the large set.

Training and testing the classification algorithms:

The dataset is further trained and tested using classification algorithms which will give us accurate results and helps us to identify the attack in a timely manner. Neural Networks are classification algorithms inspired from the understanding and emulating human brain and have been applied to diversified applications. They are structured in the form of interconnected nodes with an input layer, an output layer and one or more hidden layers.

The model has various layers where input to each node is produced by applying some non-linear function on the data. Deep [neural networks](https://www.sciencedirect.com/topics/computer-science/neural-networks) (DNNs) are improved versions of the conventional ANN with multiple layers. The DNN models are recently becoming very popular due to their excellent performance to learn not only the nonlinear input–output mapping but also the underlying structure of the input data vectors.

The DNN training involves two passes based on the error [backpropagation algorithm](https://www.sciencedirect.com/topics/engineering/backpropagation-algorithm), which are the forward pass and the backward pass. In the former one, the affine transformation and nonlinear activation are calculated layer by layer from the input to the output layer. In the later one, the derivatives of the error function with respect to individual weights are calculated in a reverse order, that is, from the output layer to the input layer.

**REFERENCES:**

1. DeepADV: A Deep Neural Network Framework for Anomaly Detection in VANETs, Tejasvi Alladi,Bhavya Gera, Ayush Agrawal, Vinay Chamola and Fei Richard Yu , IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 70, NO. 11, NOVEMBER 2021
2. DeepVCM: a Deep Learning Based Intrusion Detection Method in VANET, Yi Zeng, Meikang Qiu, Dan Zhu, Zhihao Xue, Jian Xiong, and Meiqin Liu, IEEE Intl Conference on Intelligent Data and Security (IDS).
3. Deep Learning Approach for Intelligent Intrusion Detection System, R. Vinayakumar, Mamoun Alazab, K. P. Soman , Prabaharan Poornachandran , Ameer Al-nemrat , And Sitalakshmi Venkatraman, IEEE Access ( Volume: 7)
4. Deep Neural Networks for Securing IoT Enabled Vehicular Ad-Hoc Networks, Tejasvi Alladi, Ayush Agrawal , Bhavya Gera, Vinay Chamola, Biplab Sikdar, Mohsen Guizani, CC 2021 - IEEE International Conference on Communications.
5. Attack classification using feature selection techniques: a comparative  
   study, Ankit Thakkar1 · Ritika Lohiya, Journal of Ambient Intelligence and Humanized Computing (2021) 12:1249–1266.
6. Impact of Feature Selection Techniques for Tweet Sentiment Classification , Joseph D. Prusa, Taghi M. Khoshgoftaar, David J. Dittman Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference.