A MACHINE LEARNING-BASED LIGHTWEIGHT INTRUSION DETECTION SYSTEM FOR THE INTERNET OF THINGS

RAVI KUMAR NALLI[1]

M.Tech Student (Computer Science and Technology), Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam.

ABSTRCAT *The Internet of Things (IoT) is vulnerable to various attacks, due to the presence of tiny computing devices. To enhance the security of the IoT, this paper builds a lightweight intrusion detection system (IDS) based on two machine learning techniques, namely, feature selection and feature classification. The feature selection was realized by the filter-based method, thanks to its relatively low computing cost. The feature classification algorithm for our system was identified through comparison between logistic regression (LR), naive Bayes (NB), decision tree (DT), random forest (RF), k-nearest neighbor (KNN), support vector machine (SVM) and multilayer perceptron (MLP). Finally, the DT algorithm was selected for our system, owing to its outstanding performance on several datasets. The research results provide a guide on choosing the optimal feature selection method for machine learning.*

1. INTRODUCTION

IOT devices are often deployed in a hostile and insecure environment, making them more vulnerable to different attacks [3]. Therefore, security solutions are essential to protect IoT devices from intruder attacks. An Intrusion Detection System (IDS) is a tool used to detect attacks against a system or a network by analyzing their activities and events [4]. It can act as a second line of defense which from intruders [5]. The main purpose of an IDS is to detect as many attacks as possible with an acceptable accuracy while minimizing energy consumption in resource constrained [6]. There are mainly two types of IDS, signature-based and anomaly-based IDS. A signature-based IDS also known as misuse-based IDS, detects intrusions by comparing new data with a knowledge base or signatures of known attacks.

Many researches have been recently performed in the areas of IoT and IDS to provide the best security mechanism. Sedjelmaci et al. [3] were interested to a light anomaly detection technique based on the concept of the game theory. The authors use the Nash equilibrium to predict the equilibrium state that allows the IDS agent to detect the signature of a new attack. Li et al. [7] proposed a new intrusion detection system based on the K-Nearest Neighbor (KNN) classification algorithm in a wireless sensor network. The system can detect a flood attack in the wireless sensor network. It also conducts experiments to study the effects of a flood attack. Thanigaivelan et al. [8] presented a distributed internal anomaly detection system for the Internet of Things. The main features of the system are

monitoring, ranking, isolation and reporting. Nodes monitor and note their neighbors at one hop, and if a neighbor does not maintain the required rating, the neighboring node is classified as an anomaly.

Shahid Raza [4] proposed a real-time intrusion detection system in the IoT called SVELTE. It is an IDS available in IoT that is implemented in Contiki OS. This approach only detects content spoofing attacks within the network, gulp and selective transfer attacks. Douglas et al. [9] presented an ultra-lightweight deep-packet anomaly detection approach that is possible to run on small IoT devices. The approach uses n-gram bit-patterns to model payloads and allows the n-gram size to vary by dimension.

The main objective is The Internet of Things (IoT) is vulnerable to various attacks, due to the presence of tiny computing devices. To enhance the security of the IoT, this paper builds a lightweight intrusion detection system (IDS) based on two machine learning techniques, namely, feature selection and feature classification.

The feature selection was realized by the filter-based method, thanks to its relatively low computing cost. The feature classification algorithm for our system was identified through comparison between logistic regression (LR), naive Bayes (NB), decision tree (DT), random forest (RF), k-nearest neighbor (KNN), support vector machine (SVM) and multilayer perceptron (MLP).

2. EXISTING SYSTEM

An Intrusion Detection System (IDS) is a mechanism that detects intrusions or attacks against a system or a network by analyzing the activity of the network and the system. Such intruders can be internal or external [14]:

Internal intruders are users inside the network that attempt to raise their access privileges to misuse non-authorized privileges while external intruders are users outside the target network attempting to gain unauthorized access to the network [15]. The IDS monitors the operations of a host or a network, alerting the system administrator when it detects a security violation. There are mainly three components of the IDS.

2.1 IMPLEMENTATION

The below flow chart shows the implementation of the study

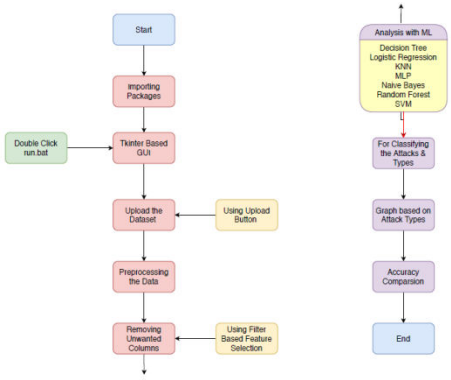


Fig 1: Flow chart

3 PROPOSED SYSTEM

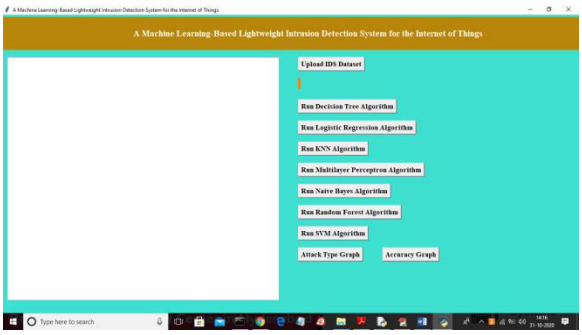
One of our main goals is that the IDS should be lightweight and comply with the

processing capabilities of the constrained nodes. Thus, according to [20], it is not possible to have an active intrusion detection agent in each node of an IoT due to the limited processing capacity and power consumption. Therefore, we have adopted a centralized IDS architecture to overhead the problem of limited capacity on the one hand and the peripheral heterogeneity issue on the other hand, where the IDS is implemented on the network layer of the IoT above the Gateway component. the activity diagram of our Lightweight Intrusion Detection System (LIDS) that consists in detecting an intrusion by observing the current behavior and comparing it to the normal behavior. If there is a deviation between the two behaviors, an alarm will be triggered. It is composed of three phases:

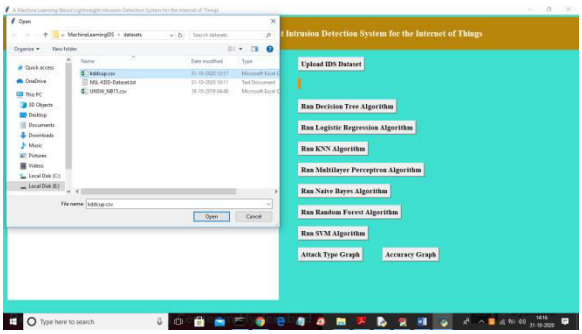
3.1 Advantages

The results of our experiments are presented in this section. Firstly, we measure the classification evaluation metrics of the three datasets with full features. Then, we reduced datasets dimension using the three popular correlation methods cited above and we evaluate each algorithm based on the best hyper parameters.

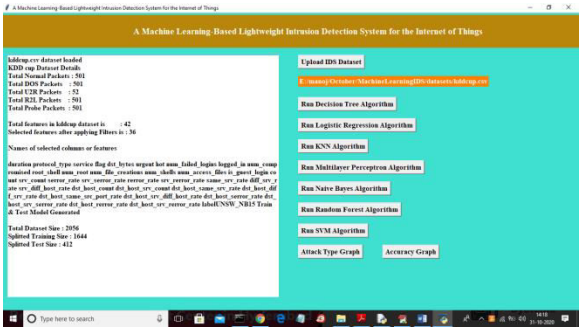
4. RESULTS AND DISCUSSIONS



In above screen click on 'Upload IDS Dataset' button and upload any one dataset

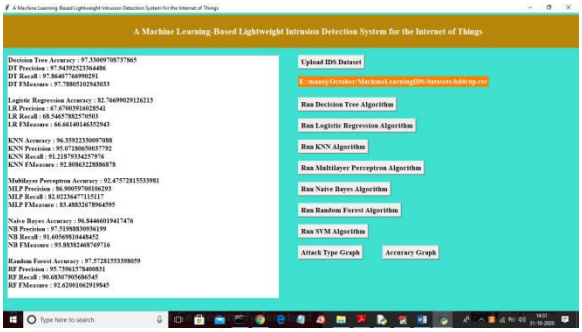


In above screen selecting and uploading 'kddcup.csv' dataset and click on 'Open' button to load dataset and then application will perform dataset preprocesing and then apply feature selection algorithm and then split dataset into train and test part.

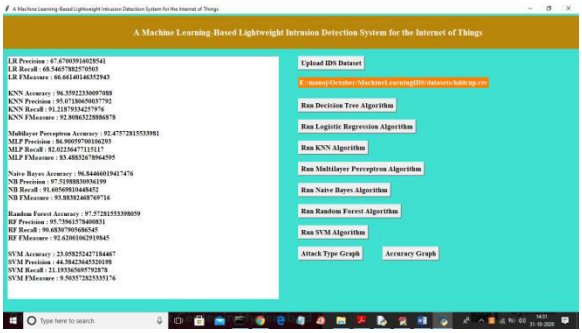


In above screen application displaying various attacks size available in dataset and then displaying total features and then displaying selected features and then

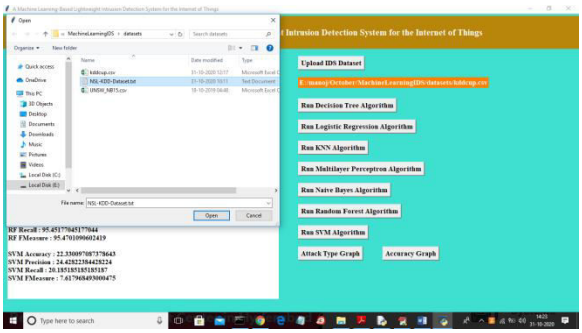
displaying names of all selected features and then displaying total dataset size and total records used for training and testing. Now both train and test data ready and now run all algorithms by clicking on each button and then calculate accuracy, precision, recall and FScore on test data.



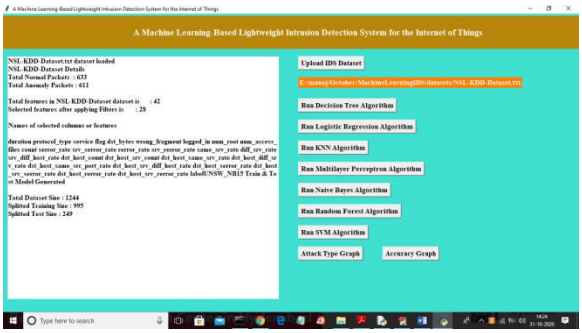
In above screen after clicking on each button we algorithm will generate model by using trained data and then classify test data to calculate accuracy and other metrics. In above screen decision tree is giving better performance. Below screen showing SVM result for KDDCUP dataset



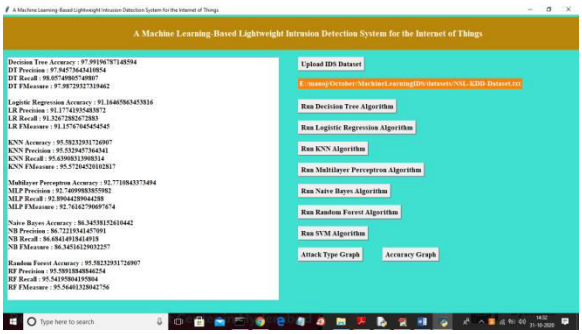
Now upload NSLKDD dataset and run all algorithms on that dataset



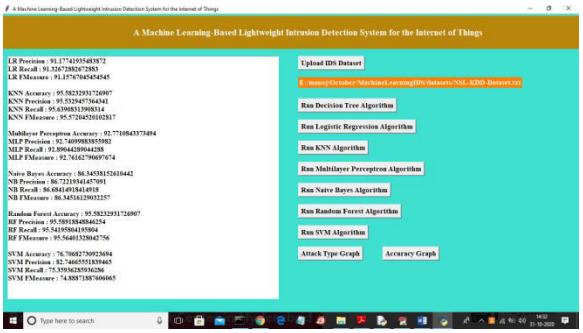
In above screen by clicking on first button uploading 'NSL-KDD-Dataset.txt' file and now click on 'Open' button to load dataset and to get below screen



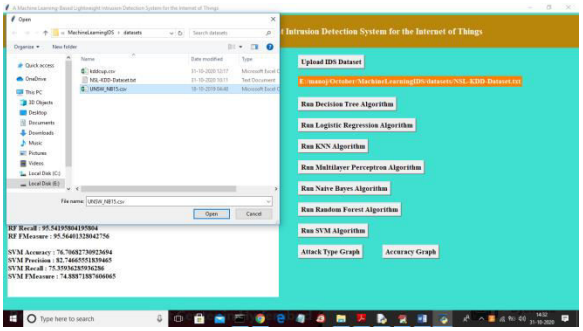
In above screen we can see all dataset details from NSL KDD data and now click on each algorithm button to calculate its accuracy



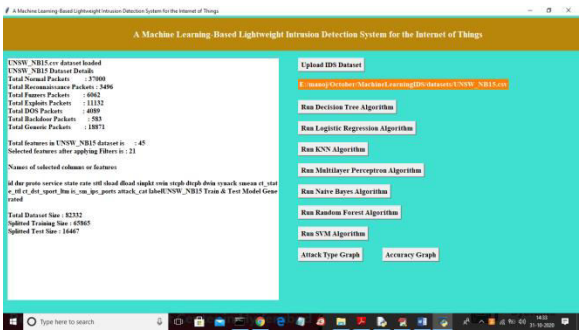
In above screen we can see algorithms accuracy for NSL KDD dataset and below is the SVM accuracy for NSL KDD



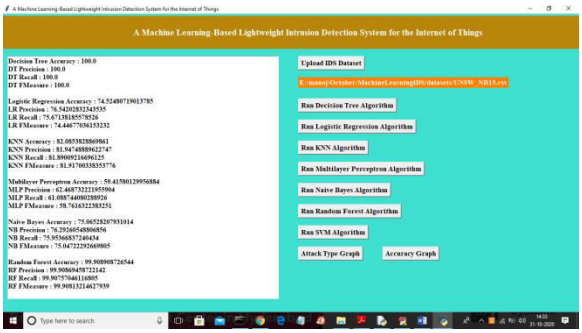
Now once again click on first button and upload UNSW dataset



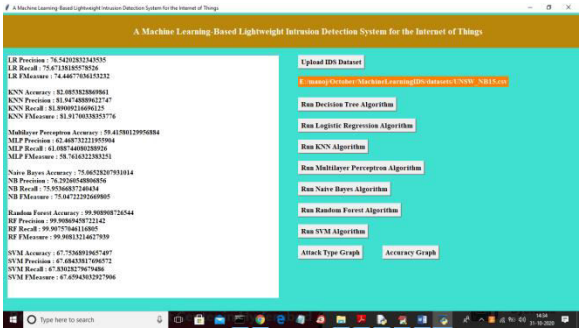
In above screen uploading third dataset called 'UNSW_NB15.csv' file and now click on 'Open' button to load dataset and to get below screen



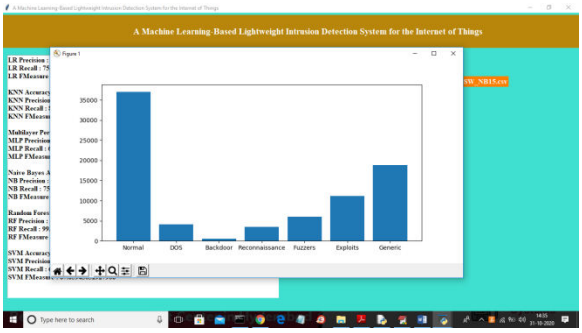
In above screen we can see all details from UNSW dataset and now both train and test data is ready and now click on each algorithm button to calculate accuracy



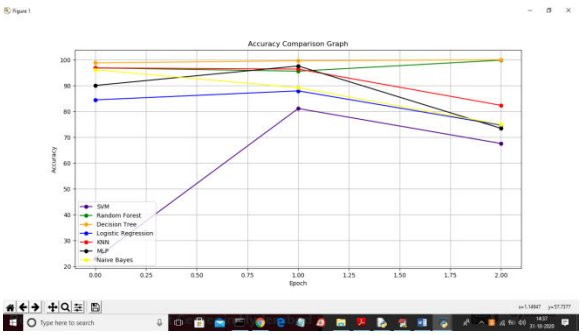
Now in below screen we can see SVM accuracy for UNSW dataset



In all datasets and in all algorithms Decision Tree and KNN is giving better results and now click on 'Attack Type Graph' button to get below graph



In above graph x-axis represents attack name and y-axis represents count of each attack and now click on 'Accuracy Graph' button to get below accuracy graph



In above graph x-axis represents 3 datasets and y-axis represents accuracy of each algorithm for that dataset. From all algorithms we can decision tree is giving good performance.

5. CONCLUSIONS

Internet of Things is increasingly used and many related applications appeared. However, the IoT is faced with a security problem that needs to be solved, while considering the constraints and challenges related to the IoT context. In this paper, we have proposed a lightweight intrusion detection model based on machine learning techniques. This model can detect new attacks and provide double protection to the IoT nodes against internal and external attacks. In order to find the best classifier model, we evaluated several machine learning classifier models using three lightweight feature selection algorithms and tried to optimize the parameters of each algorithm to get an efficient classifier model with high accuracy and precision, as well as low false negative. In the experiments, we used KDD99, NSL-KDD and UNSW-NB15 dataset to learn and evaluate our model. According to the results of our study, it is observed that DT and KNN performed better than the other algorithms; however, the KNN takes much time to classify compared to the DT algorithm. Furthermore, with the three correlation methods used to reduce datasets dimension such as PCC, SCC and KTC, the classifiers produce good performance when the threshold of the correlation coefficient is greater than 0.9; below this threshold, performances are poor and sometimes unacceptable. In the case of the datasets that relate to the extent of our study area, it is found that the performance obtained on the NSL-KDD dataset is better compared to the KDD99 and UNSW-NB15 datasets

REFERENCES

[1]Atzori, L., Iera, A., Morabito, G. (2010). The Internet ofThings: A survey. Computer Network, 54(15): 2787-2805. https://doi.org/10.1016/j.comnet.2010.05.010

[2]Weiser, M. (1991). The computer for the 21st century.Scientific American, 265(3): 94-105.

[3]Sedjelmaci, H., Senouci, S.M., Al-Bahri, M. (2016).Lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology.IEEE ICC - Mobile and Wireless NetworkingSymposium. https://doi.org/10.1109/ICC.2016.7510811

[4]Raza, S.,Wallgren, L., Voigt, T. (2013). SVELTE: Real-time intrusion detection in

the Internet of Things. Ad HocNetworks, 11(8): 2661-2674.https://doi.org/10.1016/j.adhoc.2013.04.014

[5]Anand, A., Patel, B. (2012). An overview on intrusiondetection system and types of attacks it can detectconsidering different protocols. International Journal ofAdvanced Research in Computer Science and SoftwareEngineering, 2(8): 94-98.

[6]Rajasegarar, S., Leckie, C., Palaniswami M. (2008).Anomaly detection in wireless sensor networks. IEEEWireless Communications, 15(4): 34-40.https://doi.org/10.1109/MWC.2008.4599219

[7]Li, W.C., Yi, P., Wu, Y., Pan, L., Li, J.H. (2014). A newintrusion detection system based on KNN classificationalgorithm in wireless sensor network. Journal ofElectrical and Computer Engineering, 2014: 8 pages.http://dx.doi.org/10.1155/2014/240217

[8]Thanigaivelan, N.K., Nigussie, E., Kanth, R.K., Virtanen,S., Isoaho, J. (2016). Distributed internal anomalydetection system for Internet-of-Things. 13th IEEEAnnual Consumer Communications & NetworkingConference (CCNC).https://doi.org/10.1109/CCNC.2016.7444797

[9]Summerville, D.H., Zach, K.M., Chen, Y. (2015). Ultra-lightweight deep packet anomaly detection for Internet ofThings devices. 2015 IEEE 34th InternationalPerformance Computing and CommunicationsConference (IPCCC).https://doi.org/10.1109/PCCC.2015.7410342

[10]Huang, S.H. (2003). Dimensionality reduction inautomatic knowledge acquisition: A simple greedysearch approach. IEEE Transactions on Knowledge andData Engineering, 15(6): 1364-1373.https://doi.org/10.1109/TKDE.2003.1245278

[11]Zhao, K., Ge, L. (2013). A survey on the Internet ofThings security.in Int'l Conf. on ComputationalIntelligence and Security (CIS), pp. 663-667.https://doi.org/10.1109/CIS.2013.145

[12]Leo, M., Battisti, F., Carli, M., Neri, A. (2014). Afederated architecture approach for internet of things
security. in Euro Med Telco Conference (EMTC), pp. 1-5.https://doi.org/10.1109/EMTC. 2014. 6996632