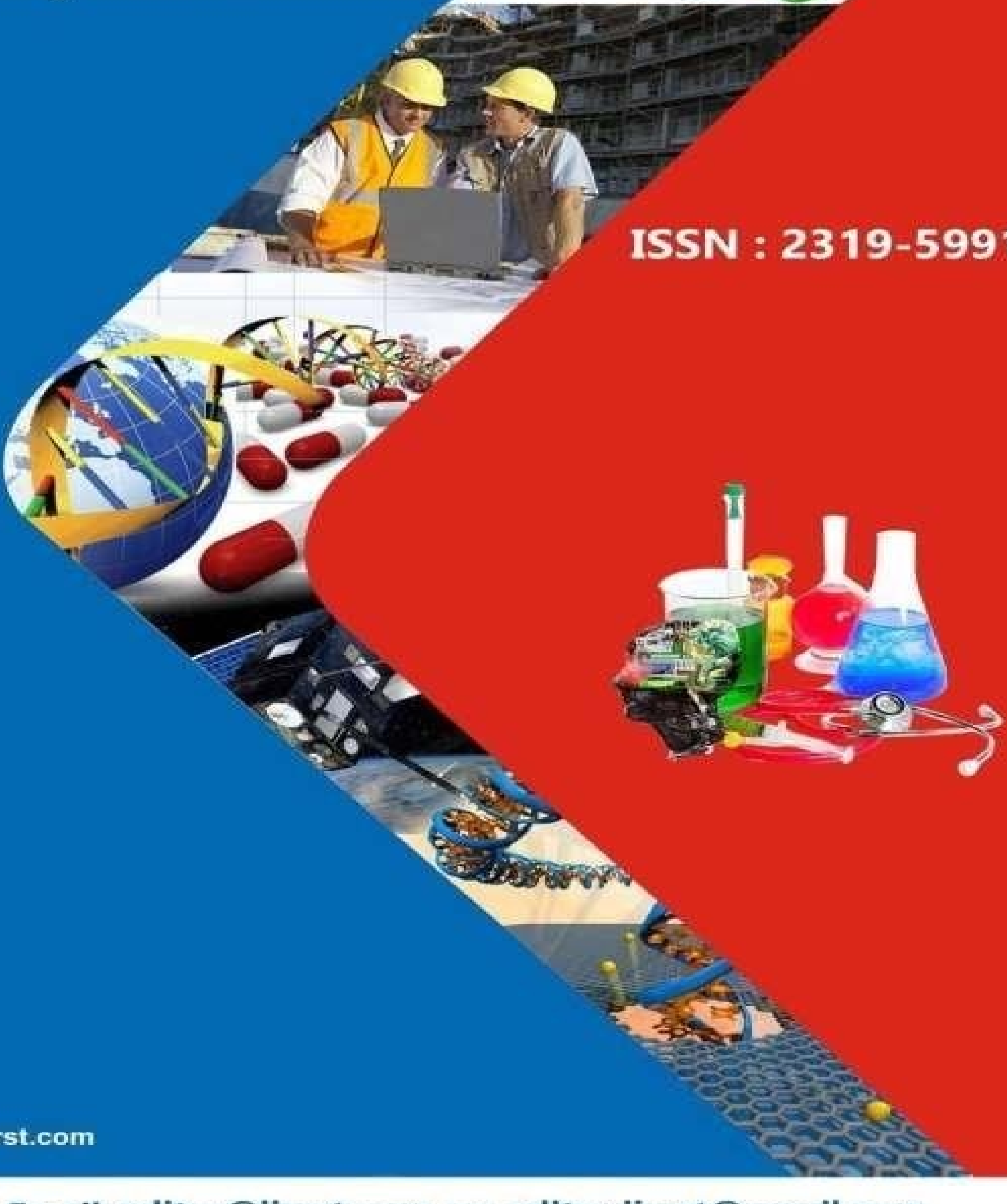
IJERST

International Journal of Engineering Research and Science & Technology



ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

https://doi.org/10.62643/ijitce.2024.v20.i3.pp165-170 ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

Anomaly Detection in Internet of Things Using Classification Techniques

[1] Ch.Venkatesh , [2] V. Agaphi Bob , [3] P. Anusha [4] P. Harinarayana
[1,2,3,4] Assistant Professors, Ramachandra College of Engineering
Eluru, Andhra Pradesh, India.

Abstract–The vulnerability of Internet of things (IoT) to attacks has become a problem. As the security concerns for these IoT devices become much more difficult,an intrusion detection system for each and every IoT device will become cumbersome as it is not a cost-effective solution. Hence, a centralized anomaly detection system is proposed in which some of the most famous classification algorithms are implementedand evaluated for the chosen data sets.

Keywords: IoT, Intrusion Detection, Classification Techniques, Machine Learning.

1. Introduction

IoT has changed the world around us making us do our works more ease.IoT is collection of devices which are connected to the internet for the data transmission and communication between the devices through a wireless network without any human help.As the devices are increasing these days the attacks on these devices are also increasing which is known as intrusion. Intrusion leads for an unauthorized person to access the data or a person to access the data in an unauthorized area. This leads to data manipulation, no data, and misuse of data and so on.Therefore, we have to protect our IoT devices from these types of attacks or intrusions. The system which can detect this type of intrusions is known as intrusion detection system.In our project we build an IDS which can detect the intrusion in IoT devices. Having an intrusion detection system at each and every node in the IoT devices is not a cost-effective solution. Hence, we are going for a centralized intrusion detection system where we have a group of IoT devices with one centralized IDS.In our project we have taken two techniques one is feature selection and the other one is feature classification.In the feature selection process, we take only those features from our data set on which the detection of attack is more dependent. In the classification process we have taken the seven most famous classification algorithms and evaluated their performance and found that suits best for intrusion detection in IoT devices.

2. Literature survey

Numerous studies have been conducted in order to secure IoT devices. SVELTE is the IoT IDS ShahidRaza suggested. It is an IDS that uses the Contiki operating system. However, this approach can only identify gulp, selective transfer, and network-wide content spoofing attacks[12]. A method of detecting light anomalies based on the idea of game theory attracted Sedjelmaci et al. [13]. The basic goal of an IDS is to identify as many attacks as possible with reasonable accuracy while consuming the least amount of energy when resources are scarce [14]. For the wireless sensor network, Li et al. [17] suggested an IDS employing a KNN classification method. However, because it can only identify a flood assault, this approach is less useful.

3. Proposed system

https://doi.org/10.62643/ijitce.2024.v20.i3.pp165-170 ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

Due to various attacks on IoT devices we need to have anIDS for these IoT devices. Therefore, we are providing an anomaly detection IDS for these IoT devices. It is a centralized IDS instead of having the intrusion detection system at each and every node which will be cost effective for small IoT devices.[9].

There are two types of detection methods of IDS, signature based and anomaly based. Anomaly-based IDS is used to detect unknown malware attacks as new malware are developed rapidly. In anomaly-based IDS there is use of machine learning to create a trustful activity model and anything coming is compared with that model and it is declared suspicious if it is not found in model. Machine learning-based method has a better-generalized property in comparison to signature-based IDS as these models can be trained according to the applications and hardware configurations. An intrusion is discovered using an IDS approach that is based on anomaly detection. It anticipates the system's typical behavior and tracks changes to that behaviour. Any activity that deviates from the usual is labeled as an intrusion [16, 17]. This method's key benefit is that it may be used to identify new threats by alerting users to any divergence from expected behaviour. However, it frequently produces a large number of false positives since a change from the expected behaviour does not always indicate an assault.Figure 1 shows the architecture of Anomaly based intrusion detection system.

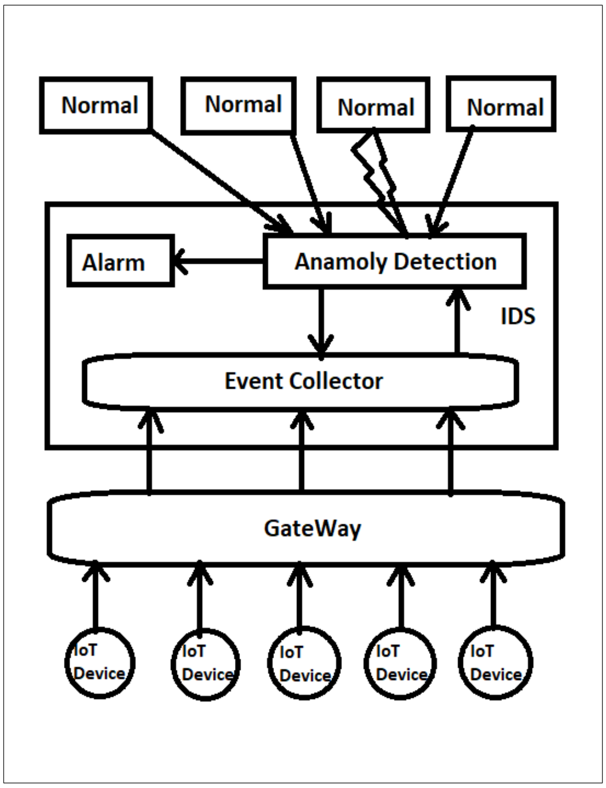


Fig 1:IDS Architecture

This Intrusion Detection System works by watching the present activity and contrasting it with the expected behaviour in order to find an intrusion. An alarm will sound if there is a difference between the two behaviours.All IoT device events are gathered and recorded by the LIDS component, Events collector in order to create the current behaviour that will be represented as a feature vector.Intruders are analysed and found during the detection phase. It is the primary element. The suggested system stops the user once an attack is detected, closes the user's session, and then notifies the administrator to take the necessary action.

To implement proposed intrusion detection system,the following steps are involved:

Upload data:

Three datasets are used, including the UNSW-NB15 dataset, the KDD Cup 99 dataset, and the NSL-KDD dataset, to develop the Anomaly based IDS.

Preprocessing:

In the preprocessing step some changes are made to the dataset and converted into a new dataset which is required for our machine learning model. In this preprocessing of the dataset, the dataset undergoes several changes such as data transformations, binarization, standardization and normalization.

Selecting features:

Here out of all the available features from the dataset we choose only those features on which the output mostly depends on. To select the features from dataset we have some feature selection models like wrapper method, embedded method and filter method. Out of all the above methods we have taken filter method, as it is more efficient and cost effective.This filter method gives a correlation value to all the features based on how the output is dependent on this feature. We choose a threshold value and select only those features whose correlation value is greater than this chosen threshold value.

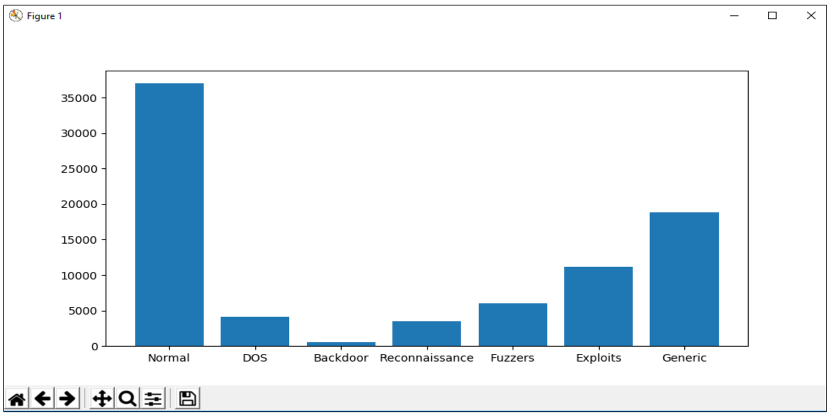
Classification techniques:

The two main phases of any machine learning work are training and testing the model. In the data preprocessing we have divided our data for training and testing. In the training phase we train our classifier with the labeled dataset. To evaluate the performance of our model we test our classifier with testing dataset. We are using the top seven well-known classification algorithms in our research, including Naive Bayes, Logistic Regression, Decision Tree, K-Nearest Neighbor, Multi-Layer Perceptron, Random Forest, and Support Vector Machine.

The dataset acquisition method is the initial step. The dataset is gathered and divided into training and testing datasets during this phase. After that, the data is cleaned through the pre-processing process, and the data dimension can be decreased through the feature selection process. Different classifier models, including Logistic Regression, Random Forest, Decision Tree, SVM, and others, were used in this study. The models are trained using the training set. Then, these models are assessed using several evaluation criteria against the testing set. Finally, three datasets will go through these steps again.

4. Results and Discussion

The common benchmark dataset with packet-based distribution appropriate for IDS testing is NSL-KDD [11]. The NSL-KDD collection includes 147,907 cases with 43 characteristics, of which 76,967 are normal instances and 70,940 are attack instances. All assaults in the collection can be categorized into one of four attack groups: DoS, Probe, R2L, and U2R. The dataset contains 43 features, of which 35 are numerical, 4 are categorical features with bi-values, 3 are categorical features with multiple values, and 1 is a feature for the class name.



https://doi.org/10.62643/ijitce.2024.v20.i3.pp165-170 ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

Fig 2: Count of each attackafter experimentation.

In above graph x-axis shows the attack type name and y-axis the count of each attack.

Here, we assess each algorithm's success in order to determine the best classifier for our particular issue. We can see from the image below that the decision tree method performs the best across all three data sets.The results of our trials are presented in this part. We start by calculating the full feature classification evaluation scores for the datasets. The best hyper parameters were used to evaluate each method.

The performance of the various classifiers using the different dimensions of the KDD99 dataset is shown below for the case of accuracy. Similarly precision and F1 score are evaluated for all other datasets.

Additionally, it can be shown that the DT classifier outperforms the competition in terms of the various metrics, with an accuracy rate of 98% across nearly all dataset dimensions and a false positive rate of less than 2% throughout.

The DT and Random forest algorithms provide the greatest performances, based on the aforementioned results. Additionally, feature selection methods occasionally give outcomes that are worse than the original features while few producing results that are similar to the original data. But most of the cases call for these strategies to function well.

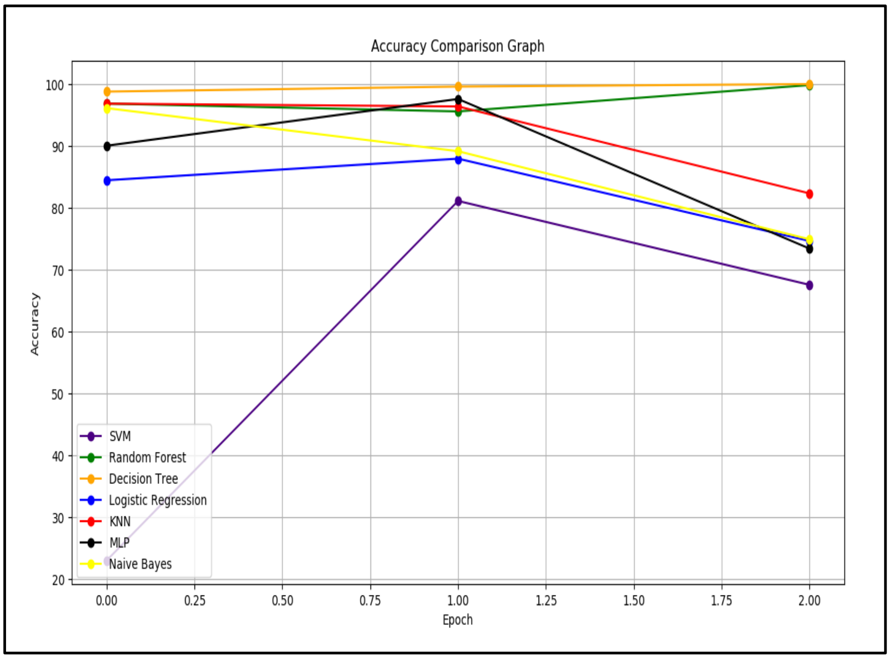


Fig 3: Accuracy graph

The performance of each and every algorithm is evaluated to find the best classifier which suites our problem.From the above graph we can observe that the decision tree algorithm is showing the best performance for all the three data sets.

https://doi.org/10.62643/ijitce.2024.v20.i3.pp165-170 ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

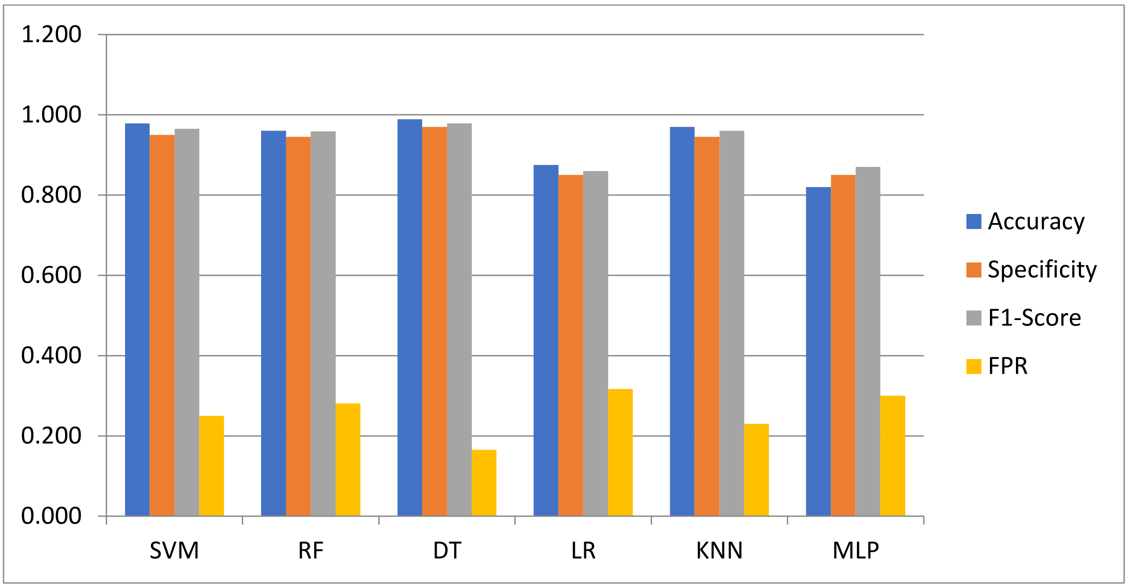


Fig 3: Performance of various algorithms

5. Conclusion

This work helps in findinga model that can detect the intrusion with more accuracy in less time. After executing various classification algorithms for the given data sets, It is observed that the DT, SVM and KNN are giving more accuracy compared to the other algorithms. The performance of DTclearly outperforms other methods . Hence, this paper proposes DT algorithm to be more suitable for intrusion detection in IoT devices. This work focuses on some of the famous algorithms. In future other algorithms can be implemented and evaluatedin real time IoT devices for intrusion detection.

REFERENCES

[1] Atzori, L., Iera, A., Morabito, G. (2010). The Internet of Things: 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] Leo, M., Battisti, F., Carli, M., Neri, A. (2014). A federated architecture approach for internet of things security. in Euro Med Telco Conference (EMTC), pp. 1- 5. https://doi.org/10.1109/ EMTC. 2014. 6996632

[4] Sherasiya, T., Upadhyay, H., Patel, H.B. (2016). A survey: Intrusion detection system for Internet of Things. International Journal of Computer Science and Engineering (IJCSE), 5(2): 91-98.

[5] KDD cup 99 Intrusion detection dataset. http://kdd.ics.uci.edu/databases /kddcup99/ kddcup.data_10_percent.gz, accessed on March 1, 2019. [6] NSL-KDDDataset, https://www.unb.ca/cic/datasets/nsl.html, accessed on March 1, 2019.

[7] Paliwal, S., Gupta, R. (2012). Denial of-service, probing & remote to user (R2L) attack detection using genetic algorithm. International Journal of Computer Applications, 60(19): 57-62. https://doi.org/10.5120/9813-4306

[8] N. ChandraSekhar Reddy, VemuriP,GovardhanA"Anemperical study on support vector machines for intrusion detection"-International Journal of Emerging Trends in Engineering Research (2019)

[9] N.Chandrasekhar Reddy "An Instruction Detection system for secure distributed local action detection and retransmission of pack" IJSC 1816-9503 2016

[10] N. Chandra Sekhar Reddy "Evaluation of PCA and K-means Algorithm for Efficient Intrusion Detection" IJAER Volume 12 Issue No:12 ISSN No:0973-4562 2017

https://doi.org/10.62643/ijitce.2024.v20.i3.pp165-170 ISSN 2319-5991 www.ijerst.com

Vol. 20, Issue 3, 2024

[11] N. Chandra Sekhar Reddy “An empirical study on Feature Extraction techniques for instruction detection system” JARDCS Volume 9,Issue 2 2017

[12] Raza, S., Wallgren, L., Voigt, T. (2013). SVELTE: Real time intrusion detection in the Internet of Things. Ad Hoc Networks, 11(8): 2661-2674.

[13] 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 Networking Symposium.https://doi.org/10.1109/ICC.2016.7510811

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

[15]Li, W.C., Yi, P., Wu, Y., Pan, L., Li, J.H. (2014). A new intrusion detection system based on KNN classification algorithm in wireless sensor network. Journal of Electrical and Computer Engineering, 2014: 8 pages. http://dx.doi.org/10.1155/2014/240217.

[16] Sherasiya, T., Upadhyay, H., Patel, H.B. (2016). A survey:IntrusiondetectionsystemforInternetofThings. International Journal of Computer Science and Engineering (IJCSE), 5(2):91-98.

[17]Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P.L., Iorkyase, E., Tachtatzis, C., Atkinson, P. (2016). Threat analysis of IoT networks using artificial neural network intrusion detection system. 2016 International Symposium on Networks, Computers and Communications (ISNCC).https://doi.org/10.1109/ISNCC.2016.7746067