ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

A New Machine Learning Based Lightweight Intrusion Detection System For The Internet Of Things

LABBA DEVI[1], DAYAM ANEETA[2], JAYAMANGALA SUDHARANI[3], GUDURI VIJAYALAKSHMI[4]

[1] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.
[2] ASST PROFESSOR DEPARTMENT OF COMPUTER SCIENCE, SIR C R REDDY COLLEGE, ELURU, INDIA.
[3] ASST PROFESSOR DEPARTMENT OF ELECTRONICS, SIR C R REDDY COLLEGE, ELURU, INDIA.
[4] ASST PROFESSOR DEPARTMENT OF ELECTRONICS, SIR C R REDDY COLLEGE, ELURU, INDIA.

ld@sircrreddycollege.ac.in[1],da@sircrreddycollege.ac.in[2], jesus4sudha4apple@gmail.com[3], gudurivijayalakshmi.crr@gmail.com[4]

Abstract

In this paper, Due to the prevalence of dispersed, low-powered computers, the IoT is open to a wide range of threats. The purpose of this paper is to improve the safety of the Internet of Things (IoT) by developing a lightweight intrusion detection system (IDS) using feature selection and feature classification, two machine learning methods. The filter-based method was used to select features because it requires less processing power than other approaches. After evaluating logistic regression (LR), naive Bayes (NB), decision tree (DT), random forest (RF), k-nearest neighbour (KNN), support vector machine (SVM), and multilayer perceptron (MLP), we settled on MLP as our system's feature classification algorithm (MLP). Given its stellar results across multiple datasets, the DT algorithm was ultimately chosen for implementation in our system. The findings of the study can be used as a reference when deciding which feature selection technique to employ when engaging in machine learning.

Keywords: Internet of Things (IoT), logistic regression (LR), naive Bayes (NB), decision tree (DT), random forest (RF), k-nearest neighbour (KNN),

I. INTRODUCTION

IoT devices are prime candidates for attack due to the fact that they are often deployed in environments that are not only unsecure but also potentially hazardous [3]. Implementing security precautions is absolutely necessary if one wishes to stop criminals from gaining access to Internet of Things devices. An Intrusion Detection System, also known as an IDS, is a type of security software that monitors a computer network or another type of system for unusual activity with the goal of locating potential intrusion attempts [4]. It is efficient enough to be utilised as a secondary line of defence against possible attackers [5]. When there are limited resources available, the primary objective of an intrusion detection system (IDS) is to identify as many attacks as possible while maintaining an acceptable level of accuracy and using as little energy as possible [6]. In general, intrusion detection systems (IDS) can be broken down into two categories: those that depend on signatures, and those that depend on detecting anomalies. Signature-based intrusion detection systems, also known as misuse-based intrusion detection systems, compare newly collected data to previously stored signatures of known attacks in order to detect intrusions. This method is useful in recognising common dangers; however, it is frequently unable to recognise less common dangers. An anomaly-based intrusion detection system (also known as IDS) compares normal behaviour to actual events in order to identify deviations from the norm.

In recent years, there has been a lot of work done in the fields of IoT and IDS to develop the most effective security mechanism possible. It was of particular interest to Sedjelmaci et al. [3] to use a method that was inspired by game theory in order to detect lighting anomalies. [Citation needed] This article makes use of the Nash equilibrium to predict the equilibrium state that enables an IDS agent to recognise the signature of a novel attack. This is accomplished through the use of predictive modelling. In order to protect wireless sensor networks from potential vulnerabilities, K-Nearest Neighbor (KNN) is a classification algorithm that was suggested to be used in a brand new intrusion detection system that was proposed by Li et al. [7]. The system will be able to detect a flood attack on the wireless sensor network in the event that it occurs. In addition to this, it acts out hypothetical

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

floods in order to gain a better understanding of the aftermath of such events. Thanigaivelan et al. [8] presented an article in which they discussed a distributed internal anomaly detection system for the Internet of Things. Monitoring, ranking, isolating, and reporting are the primary responsibilities of this system. Nodes make observations and keep records about their neighbours at one hop; if a neighbour is unable to maintain the required rating, the node in question is labelled as an outlier. For the Internet of Things, Shahid Raza [4] proposed a system for real-time intrusion detection that he called SVELTE. The Contiki operating system incorporates an Internet of Things intrusion detection system (IDS) as part of its core functionality. The only types of attacks that can be identified using this method are in-network content spoofing, gulp, and selective transfer attacks. An extremely lightweight deep-packet anomaly detection method that can be implemented on low-powered Internet of Things devices was presented by Douglas et al. [9]. This method can be found in their paper. The method utilises n-gram bit patterns to model payloads, with the size of the n-gram being independent of the dimensions being modelled.

II. METHODOLOGY

Internet Of Things

The internet of things, or IoT, is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

A thing in the internet of things can be a person with a heart monitor implant, a farm animal with a biochip transponder, an automobile that has built-in sensors to alert the driver when tire pressure is low or any other natural or man-made object that can be assigned an Internet Protocol (IP) address and is able to transfer data over a network.

Increasingly, organizations in a variety of industries are using IoT to operate more efficiently, better understand customers to deliver enhanced customer service, improve decision-making and increase the value of the business.

Working:

An IoT ecosystem consists of web-enabled smart devices that use embedded systems, such as processors, sensors and communication hardware, to collect, send and act on data they acquire from their environments. IoT devices share the sensor data they collect by connecting to an IoT gateway or other edge device where data is either sent to the cloud to be analyzed or analyzed locally. Sometimes, these devices communicate with other related devices and act on the information they get from one another. The devices do most of the work without human intervention, although people can interact with the devices -- for instance, to set them up, give them instructions or access the data.

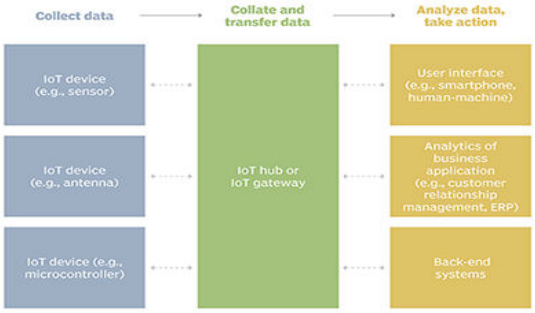


Figure 1: IoT System

Importance of IoT:

The internet of things helps people live and work smarter, as well as gain complete control over their lives. In addition to offering smart devices to automate homes, IoT is essential to business. IoT provides businesses with a real-time look into how their systems really work, delivering insights into everything from the performance of machines to supply chain and logistics operations.

IoT enables companies to automate processes and reduce labor costs. It also cuts down on waste and improves service delivery, making it less expensive to manufacture and deliver goods, as well as offering transparency into customer transactions.

As such, IoT is one of the most important technologies of everyday life, and it will continue to pick up steam as more businesses realize the potential of connected devices to keep them competitive.

Recurrent Neural Networks (RNNs)

Because RNNs contain memory, data can be kept indefinitely throughout the network. Check out the image that has been provided for your viewing pleasure. One RNN cell is seen on the left side of the picture; it takes an input value, x, and processes it in a hidden cell, h, before returning an output value, o. In addition to its linear structure, the hidden layer also has a loop that enables information to be passed on to the following layer. Like RNN may be thought of as a collection of

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

subnetworks that are quite similar to one another. The right side of the picture is what we would obtain if we unfolded the left side. Each input cell receives data as well as the output from the cell before it. For instance, the inputs to the second network node are Xt and Ot-1, the result of the node before it.

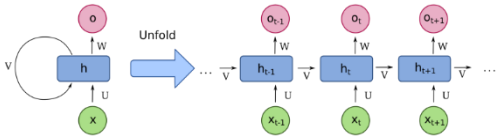


Figure 2: Recurrent Neural Networks

The chain-like topology of recurrent neural networks demonstrates their close relationship to sequences. So, if we want to translate some text from one language to another, we may provide individual words as the input. The RNN allows for a better understanding of a sentence's context by conveying information from one cell to the next.

Long short-term memory (LSTM)

By applying specific formulas to the input data, the RNN creates a new version of the data whenever new information is introduced to the network. That way, there won't be any kind of evaluation of relative significance. The LSTM's cell-state mechanism ensures that the problem is always at the right place. Therefore, LSTMs are able to retain such thoughts as "this data is necessary, and this is not that much." This is a typical LSTM's structure.

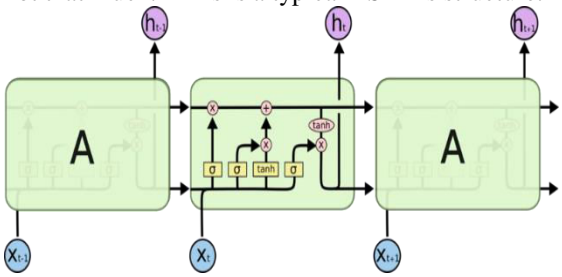


Figure 3: Long short-term memory

Different operations are performed on the inputs by the cells labelled Xt-1, X, and Xt+1. Each cell first imagines the state and output of the preceding cell, then it constructs its own cell state and output, which it would then broadcast to the very next cell, and so on.

Similar to RNNs, LSTMs lose track of important details as input size or complexity grows. Because the cell must have forgotten or the cell state must have been overwritten, it becomes impossible to maintain the context of a word that is far removed from the location where it is needed. As it progresses, the value of words would decline.

Another issue is that neither RNNs nor LSTMs are capable of doing the work in parallel. That example, in a model for translating across languages, data are sent to the network one word at a time. Given that each word is sent as input to the network, the amount of time required to compute anything would be proportional to the length of the text as a whole.

Convolutional Neural Networks (CNNs)

To aid with parallelization, local dependencies, and positional distance, we may use Convolutional Neural Networks (CNNs).

The convolutional neural networks (CNNs) Wavenet and Bytenet are commonly utilised for sequence transduction.

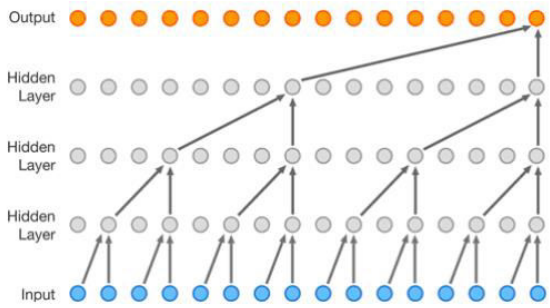


Figure 4: Convolutional Neural Networks

Because each word on the input does not always rely on the preceding words to be translated, CNNs may process them simultaneously. Furthermore, for a CNN, the "distance" between the output word and any input is on the order of log(N), where N is the number of words in the training set and the height of the tree created from the output to the input is the distance (you can see it on the GIF above). That's a huge improvement over RNNs, whose output is often N times as far from the input as the input itself.

PROPOSED SYSTEM

In order to achieve one of our primary goals, which is to make the IDS as lightweight as is practically possible while still satisfying the requirements for the processing capabilities of the constrained nodes, one of our priorities is to. According to [20], it is not possible to have an active intrusion detection agent in each node of the network due to the limited processing capacity and power consumption of individual nodes in an Internet of Things network. This is one of the reasons why it is not possible to have an active intrusion detection agent. Because of this, in order to circumvent the problem of limited capacity on the one hand and

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

the problem of peripheral heterogeneity on the other, we have opted to utilise a centralised IDS architecture. The IDS implementation is placed in this architecture on the network layer of the IoT, which is located above the Gateway component in the hierarchy. the activity diagram of our Lightweight Intrusion Detection System (LIDS), which identifies an intrusion by monitoring the behaviour that is occurring at the moment and contrasting it to the behaviour that is typical for the system. The purpose of making this comparison is to establish whether or not there has been an intrusion into the system. A warning will be given in the event that there is a disparity between the two behaviour patterns that have been established.

Advantages:

In this section, the outcomes of our experiments are detailed for your perusal. In the first step of this process, we evaluate the three datasets with their full features by measuring the classification evaluation metrics. After that, we used the aforementioned three popular correlation methods to reduce the dimensions of the datasets, and then we evaluated each algorithm based on which one had the best hyper parameters.

III. RESULTS & DISCUSSION

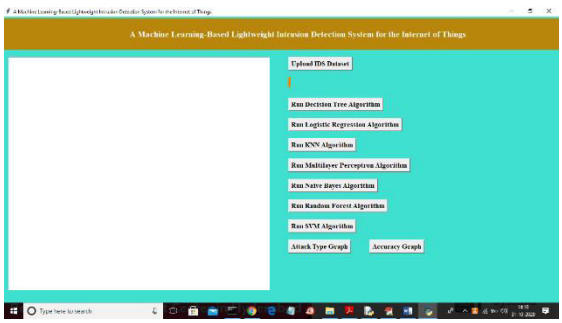


Figure 5: Upload IDS Dataset

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

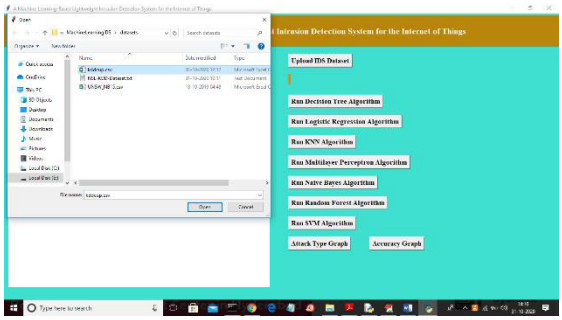


Figure 6: click on 'Open' button to load dataset.

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.

Figure 7: displaying various attacks size available in dataset.

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.

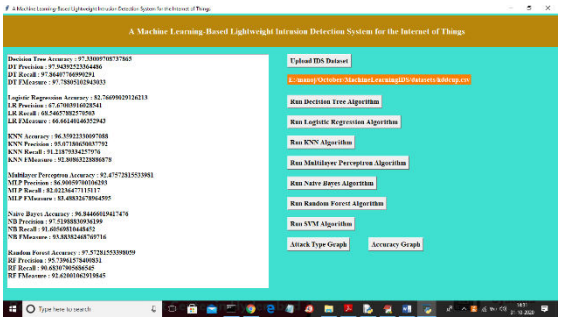


Figure 8: after clicking on each button, we algorithm will generate model by using trained data

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

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

Figure 9: upload NSLKDD dataset

Now upload NSLKDD dataset and run all algorithms on that dataset.

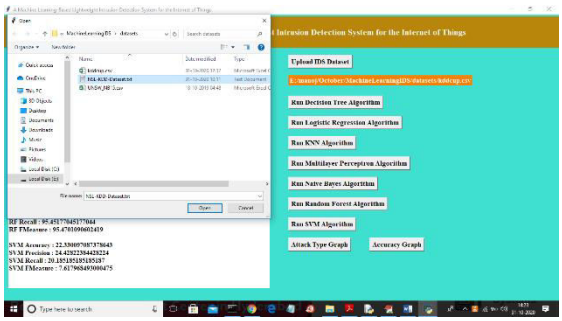


Figure 10: uploading 'NSL-KDD-Dataset.txt' file

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

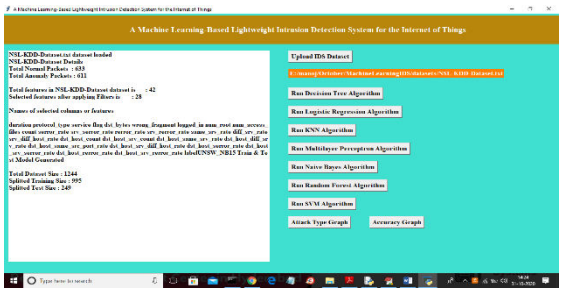


Figure 11: all dataset details from NSL KDD data

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

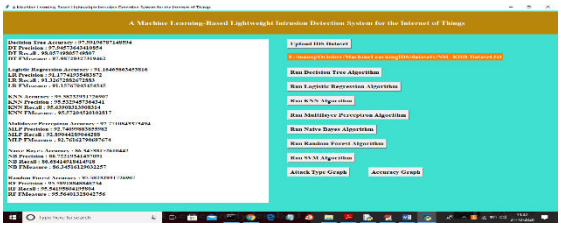


Figure 12: algorithms accuracy for NSL KDD dataset

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

Figure 13: upload UNSW dataset

Now once again click on first button and upload UNSW dataset

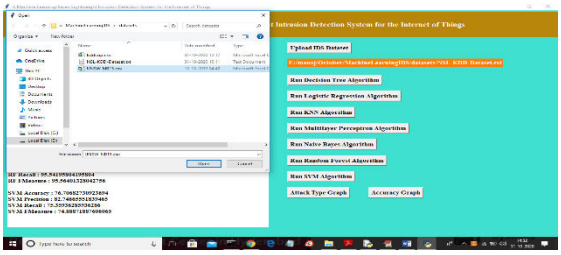
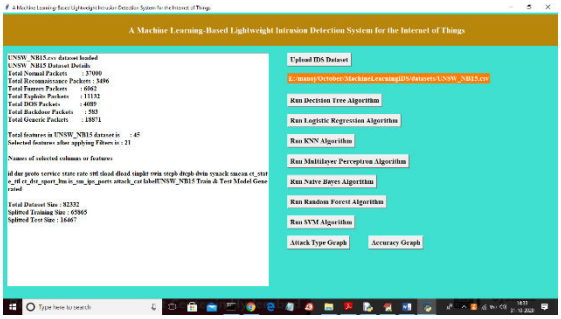


Figure 14: click on 'Open' button to load dataset

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



ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

Figure 15: all details from UNSW dataset

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

Figure 16: SVM accuracy for UNSW dataset

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

Figure 17: click on 'Attack Type Graph'

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

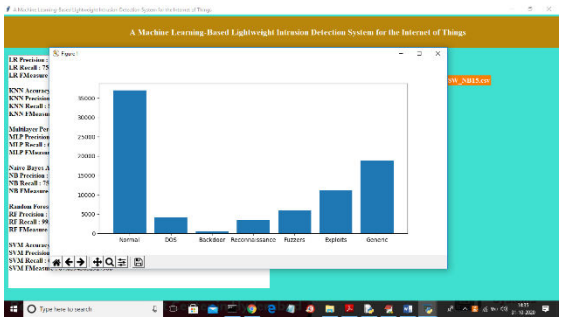


Figure 18: Accuracy Graph

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

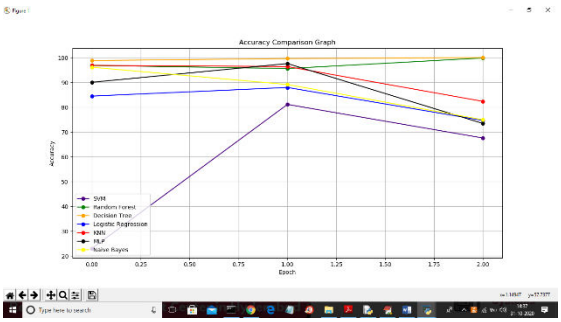


Figure 19: Accuracy Graph

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.

IV. CONCLUSION

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.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

REFERENCE

[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

[13]Oh, D., Kim, D., Ro, W.W. (2014). A malicious patterndetection engine for embedded security systems in theInternet of Things. Sensors, 14(12): 24188-24211.https://dx.doi.org/10.3390/s141224188

[14]Sherasiya, T., Upadhyay, H., Patel, H.B. (2016). Asurvey: Intrusion detection system for Internet of Things.International Journal of Computer Science andEngineering (IJCSE), 5(2): 91-98.

[15]Zarpelão, B.B., Miani, R.S., de Alvarenga, S.C. (2017).A survey of intrusion detection in Internet of Things.Journal of Network and Computer Applications, 84(C):25-37. http://dx.doi.org/10.1016/j.jnca.2017.02.009

[16]Alrajeh, N.A., Khan, S., Shams, B. (2013). Intrusiondetection systems in wireless sensor networks: A review.International Journal of Distributed Sensor Networks,2013: 7 pages. https://doi.org/10.1155/2013/167575

[17]Liao, H.J., Richard Lin, C.H., Lin, Y.C., Tung, K.Y.(2013). Review intrusion detection system: Acomprehensive review. Journal of Network andComputer Applications, 36(1): 16-24.https://doi.org/10.1016/j.jnca.2012.09.004

[18]Maharaj, N., Khanna, P. (2014). A comparative analysisof different classification techniques for intrusiondetection system. International Journal of ComputerApplications, 95(17): 22-26.http://dx.doi.org/10.5120/16687-6806

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

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 12, Iss 1, Jan 2023

[20]Roman, R., Zhou, J.Y., Lopez, J. (2006). Applyingintrusion detection systems to wireless sensor networks.In IEEE Consumer Communications & NetworkingConference (CCNC 2006), pp. 640-644.https://doi.org/10.1109/CCNC.2006.1593102

[21] KDD cup 99 Intrusion detection dataset. http://kdd.ics.uci.edu/databases /kddcup99/ kddcup.data_10_percent.gz, accessed on March 1, 2019.

[22]NSLKDDDataset,https://www.unb.ca/cic/datasets/nsl.html, accessed onMarch 1, 2019.