International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 7, pp: 303-308 www.ijaem.net ISSN: 2395-5252

"A Review on Intrusion Detection System for Internet of Things"

Poornima Dwivedi, Shatendra Dubey, Anurag Shrivastava
MTech Scholar Department of CSE NIIST Bhopal
Assistant Professor Department of CSE NIIST Bhopal
HOD Department of CSE NIIST Bhopal

Date of Submission: 25-09-2020 Date of Acceptance: 05-10-2020

ABSTRACT— The Internet of Things (IoT) combines hundreds of millions of devices which are capable of interaction with each other with minimum user interaction. Now internet of things is become the fastest-growing areas in of computing; however, the reality is that in the extremely hostile environment of the internet, IoT is vulnerable to numerous types of cyber attacks. To resolve this, Intrusion detection in the internet of things (IOT) is a rising concern practical countermeasures need to be established to secure IoT networks, such as network intrusion detection. Since IoT devices have low storage capacity and low processing power, traditional high-end security solutions to protect an IoT system are not appropriate. Also, IoT devices are now connected without human intervention for longer periods. This implies that intelligent network-based security solutions like machine learning solutions must be developed. This work proposed a new approach for classification of cyber attack. Machine learning technique also used for this system. this work talks about classification of abnormal activity, high accuracy and detection rate with low false alarm.

Keywords— Internet of Things, IDS, Classification, Machine Learning, IoT, Detection rate

I. INTRODUCTION

The Internet of Things (IoT) [2] is a network of sensing devices with limited resources and capable of wired/wireless communications with cloud services. IoT devices are being increasingly targeted by attackers using malware as they are easier to infect than conventional computers. This is due to several reasons [2] such as presence of legacy devices with no security updates, low priority given to security within the development cycle, weak login credentials, etc. This research aims to find an effective solution to security issues faced by the network environment of Internet of Things. This research will be used to develop an intrusion detection system that can detect complex and changeable Internet of things attacks, and can intelligently cope with sudden intrusions. It is also intended that the research will try to improve the performance of the system by optimizing the algorithm with regards to parameter weight and learning rate. The scope of this research will be finding a more efficient intrusion detection system for IoT devices. This work will focus on threats of IoT devices, drawbacks of existing IDS for IoT devices and usable technologies that can be used to improve IDS for IoT devices.

Use of internet and IOT devices rapidly increasing attacks are coming from the network .This new intrusion detection system will therefore focus on detection rate and false alarm rate of the system. a new Intrusion Detection System based on Machine Learning Technology for the Internet of Things. a novel model is presented for anomaly based intrusion detection of IoT, which is developed on a dimension reduction method and a classification model. Dealing with issues such as accuracy and error rate is considered with the following unique characteristics of the proposed model: Low computing complexity which is achieved by dimension reduction technique; Large dataset problem which is solved by classification models which are designed use machine learning algorithms; High detection rates can be achieved without needing large training sets and dealing with a large amount of data. Machine learning techniques are used to learn from these data to make the device or thing intelligent. The main challenges and trends of the machine learning techniques in deriving the knowledge for the IoT community to make the devices more automated. The security of Internet of things is shown in figure 1.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 7, pp: 303-308 www.ijaem.net ISSN: 2395-5252

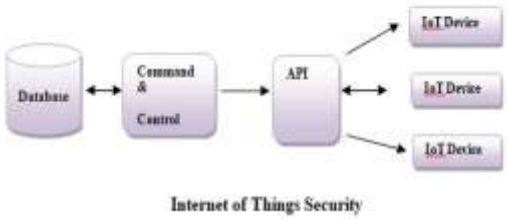

Figure : 1 IoT Security

The rest of paper is organized as follows: In section 2 we review the work related to IDS for IoT. In section 3 we describe the proposed framework for classification. Section 4 gives the detail about the dataset for IoT. and finally section 5 conclude this paper.

1.1 Intrusion Detection System

Intrusion Detection System (IDS) is an active process or device that analyzes system and network activity for unauthorized activity [2]. An ID is hardware or software or a combination of both which is used to monitor a system or network of systems against any malicious or unauthorized activities [2]. Intrusion Detection Systems (IDSs) are used to improve network security. An ID improves the security of the network by identifying, assessing, and reporting unauthorized network activities. IDS are categorized into two classes: network-based and host-based. Network based Intrusion Detection Systems analyses network packets retrieved from the network. Host-based Intrusion Detection System analyses system calls generated by individual hosts [2].The data flows through a network is very large and it is difficult to analyze and detect the attacks using traditional methods. Today we have number of Machine learning techniques available which are very useful for analyzing the data and detecting the attacks. In this paper we have used various machine learning techniques for network intrusion detection [2].

1.2 Data Mining

The iterative and interactive process of discovering valid, novel, useful, and understandable knowledge (patterns, models, rules etc) in Massive databases. Valid: generalize to the future, Novel: what we don't know Useful: be able to take some action, Understandable: leading to insight Iterative: takes multiple passes, Interactive: human in the loop. Researchers identify two fundamental goals of data mining: prediction and description. Prediction makes use of existing variables in the database in order to predict unknown or future values of interest and descriptive focus on finding patterns describing the data.

Data mining techniques: Two data mining techniques are discussed here:

Classification : Classification in data mining is data analysis task means finding rules that partition the data in to disjoint groups.

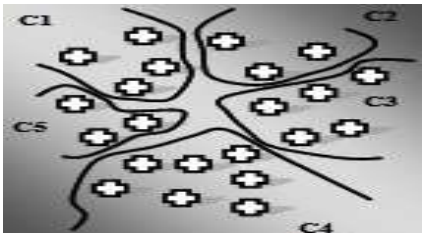

Figure : 2 Classification

Clustering: Clustering in data mining is grouping data into different groups, so that data in each group share similar trends & patterns.

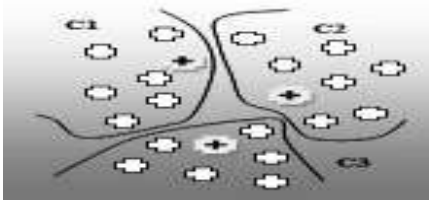

Figure : 3 Clustering

1.3 Machine learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can teach themselves to grow and change when exposed to new data. Machine learning techniques have ability to implement a system that can learn from data. For example, a machine learning system could be trained on incoming packets to learn to distinguish between intrusive and normal packet. After learning, it can then be used to classify new incoming packets into intrusive and normal packets.[23] In machine learning, computer algorithms (learners) attempt to automatically distill knowledge from example data. This knowledge can be used to make predictions about novel data in the future and to provide insight into the nature of the target concepts applied to the research at hand, this means that a computer would learn to classify alerts into incidents and non-incidents task. A possible performance measure (P)

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 7, pp: 303-308 www.ijaem.net ISSN: 2395-5252

for this task would be the Accuracy with which the machine learning program classifies the instances correctly. Machine learning often included in the category of predictive analytics as it helps to predict the future analysis.

Machine Learning Algorithms:-

K-NN Algorithms: K-Nearest Neighbors (K-NN)

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already in the beginning of 1970's as a non-parametric technique. KNN is a non parametric lazy learning algorithm. The k-Nearest Neighbor algorithm is based on learning by analogy. The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms.

Genetic Algorithm :

Genetic Algorithms, first proposed by Holland in 1975,are a class of computational models that mimic natural evolution to solve problems in wide variety of domains. Genetic algorithms are particularly it is suitable for solving complex optimization problems and for applications that require adaptive problem-solving strategies .It is based on the mechanics of natural genetics, ie.,oprations existing in nature. A genetic algorithm operates on a set of individual elements (the population) and there is a set of biologically inspired operators that can change these individuals .In computing terms, genetic algorithms map strings of numbers to each potential solution .Each solution becomes an individual in the population ,and each string becomes a representation of an individual. There should be a way to derive each individual from its string representation. The genetic algorithm then manipulates the most promising string in its search for an improved solution. The algorithm operates through a simple cycle: Creation of a population of strings, Evaluation of each string, Selection of the best strings, Genetic manipulation to create a new population of strings.

Decision Tree:-

A decision tree is a classification scheme which generates a tree and a set of rules, representing the model of deferent classes, from a given data set. The set of records available for developing classification methods is generally divided into two disjoint subsets –a training set and a test set. Attributes whose domain is numerical are called the numerical attributes, and the attributes whose domain is not numerical are called the categorical attributes. There is one distinguished attribute called the class label.

II. LITERATURE REVIEW

This research [1] involve the design of a novel intrusion detection system and the implementation and evaluation of its analysis model. This new intrusion detection system uses a hybrid placement strategy based on a multi-agent system. The new system consists of a data collection module, a data management module, an analysis module and a response module. For the implementation of the analysis module, this research work applies a deep neural network algorithm for intrusion detection. The authors results demonstrate the efficiency of deep learning algorithms for detecting attacks from the transport layer. Compared with traditional detection methods used in IDSs, the analysis indicates that deep learning algorithms are more suitable for intrusion detection in an IoT network environment.

In the research [2] authors present EDIMA, a distributed modular solution which used towards the detection of IoT malware network activity in large-scale networks (e.g. ISP, enterprise networks) during the scanning/infecting phase rather than during an attack. EDIMA employs machine learning algorithms for edge devices' traffic classification, a packet traffic feature vector database, a policy module and an optional packet sub-sampling module. We evaluate the classification performance of EDIMA through test bed experiments and present the results obtained.

Authors [3] extract the characteristics of the network traffic generated during the Internet connection, then use the information gain algorithm to select the discriminant classification features, establish the classifier by Bayesian model updating method which is an improved algorithm based on Bayesian theory, and compare with other machine learning classifiers such as k-nearest neighbor (KNN), SVM and J48, improved algorithm has good performance on validity, accuracy, efficiency and strong practicability. this experimental results and comparison, authors gives two conclusions. Firstly, the malware detection method based on traffic characteristics analysis is feasible in IoT services, and can achieve higher accuracy with few significant features.

Authors [4] proposed a model uses Principal Component Analysis (PCA) to reduce dimensions of dataset from a large number of features to a small number. To develop a classifier,

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 7, pp: 303-308 www.ijaem.net ISSN: 2395-5252

softmax regression and k-nearest neighbour algorithms are applied and compared. Experimental results using KDD Cup 99 Data Set show that our proposed model performs optimally in labelling benign behaviours and identifying malicious behaviours. The computing complexity and time performance approve that the model can be used to detect intrusions in IoT. Authors proposed a novel model which could help detect intrusion in IoT network layer. The model consists of a dimension reduction model - PCA and a classifier which is based on softmax regression. After the steps of reducing dimensions of the data and classifying types of behaviours that data represented, the selected performance indicators show the reduced-dimension data contains a slightly less useful information that helps to distinguish the types of attacks, but the noise information is significantly reduced compared to the original one.

Authors [5] present the performance of several machine learning models have been compared to predict attacks and anomalies on the IOT system accurately. The machine learning algorithms that have been used here are logistic regression, support vector machine, decision tree, random forest and artificial neural network. The evaluation metrics used in the comparison of performance are accuracy, precision and recall. This system obtained 99.4% test accuracy for decision tree, random forest and ANN.

Author [6] present three new Intrusion Detection Systems (IDSs) for IoT: 1) Kmeans clustering unsupervised learning based IDS; 2) decision tree based supervised IDS; and 3) a hybrid two stage IDS that combines K-means and decision tree learning approaches. To the best of our knowledge, these are the first machine learning based IDSs (together called as ML-IDS) for IoT. All the three IDS are centralized and scalable approaches. The K-means approach achieves 70-93% detection rate for varying sizes of random IoT networks. Decision tree based IDS achieves 71-80% detection rate and the hybrid approach attains 71-75% detection rate for the same network sizes. Although the hybrid IDS obtains lower detection rate, it is more accurate than the other two approaches. The hybrid approach eliminates the false positives significantly, while the other two IDS suffer from a higher number of false positives. Similar results are also obtained for regular mesh, star and ring topologies of IoT networks, each comprising 16 nodes. This three machine learning based centralized IDS are proposed for RPL networks in IoT. To the best of our knowledge, this is the first time machine learning has been used to develop IDS for IoT.

III. PROPOSED FRAMEWORK FOR CLASSIFICATION:

The Proposed Framework which employs simple Classification model based on machine learning (ML) techniques. The input IoT dataset is suitably processed and converted into a suitable format. The preprocessing techniques are iteratively applied in the next step, and got the cleaned dataset. Now applying the feature selection techniques for select beast features. After preprocessing and feature selection done we apply the machine learning algorithms . Here in the proposed framework, this systematic approach are applied to the model for better accuracy of classifier. At last the model is deployed and tested on test data and calculate the result.

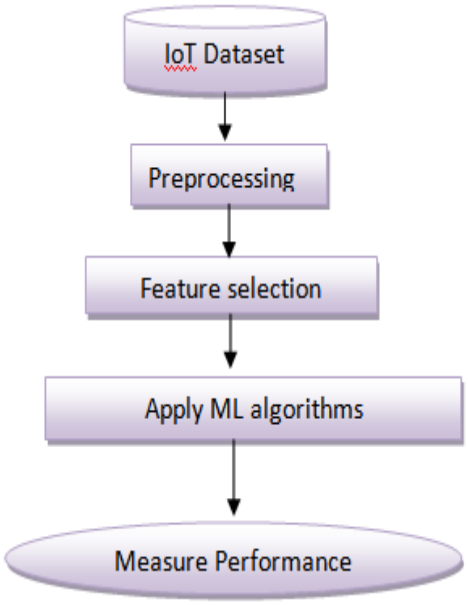

Figure 4: Proposed model for Classification

1. Dataset Description : The BoT-IoT Dataset

The BoT-IoT dataset was created by designing a realistic network environment in the Cyber Range Lab of The center of UNSW Canberra Cyber. The environment incorporates a combination of normal and botnet traffic. The dataset's source files are provided in different formats, including the original pcap files, the generated argu's files and csv files. The files were separated, based on attack category and subcategory, to better assist in labeling process. The captured pcap files are 69.3 GB in size, with more than 72.000.000 records. The extracted flow

traffic, in csv format is 16.7 GB in size. The dataset includes DDoS, DoS, OS and Service Scan, Key logging and Data ex filtration attacks, with the DDoS and DoS attacks further organized, based on the protocol used. To ease the handling of the dataset, we extracted 5% of the original dataset via the use of select MySQL queries. The extracted 5%, is comprised of 4 files of approximately 1.07 GB total size, and about 3 million records.[19]

2. Expected outcomes of the proposed work :

The Expected outcome of the proposed work is to proposed the Intrusion detection model for Internet of things and improve the IDS performance with the help of machine learning classifier accuracy. We proposed classification framework model with machine learning algorithms. we hope this proposed architectural framework gives better result. we compare the result with other classifiers. The parameter of the IDS is accuracy, detection rate, false alarm rate and error rate.

IV. CONCLUSION

In this paper machine learning technique have been proposed for the classification of IoT data. Intrusion detection system for IoT is the most important tool for detection of attack. now the use of internet of things (IoT) are increasing rapidly it is necessary to build a more effective intrusion detection model. In this work classification framework model is proposed with the Machine Learning techniques. The result evaluated in terms of accuracy, detection rate, false alarm rate and error rate. The purpose of this proposed method efficiently classify abnormal and normal data by using very large data set and detect intrusions even in large datasets with short training and testing times. Most importantly when using this method redundant information, complexity with abnormal behaviors are reduced. With proposed work machine learning classification algorithms gives high accuracy for various type of attacks. The proposed method improve the intrusion detection system performance using machine learning technique. The proposed system gives better performance in terms of high detection rate, low false alarm rate, less training and testing time, and high accuracy which is used for internet of things dataset.

REFERENCES

[1]. Chao Liang1, Bharanidharan Shanmugam1, Sami Azam1, Mirjam Jonkman1, Friso De Boer1, Ganthan Narayansamy2 "Intrusion Detection System for Internet of Things based on a Machine Learning approach" 978-1-5386-9353-7/19/$31.00 ©2019 IEEE

[2]. Ayush Kumar and Teng Joon Lim "EDIMA: Early Detection of IoT Malware Network Activity Using Machine Learning Technique" 2019 IEEE 5th World Forum on Internet of Things (WF-IoT)

[3]. Fei Wu, Limin Xiao, Jinbin Zhu "Bayesian Model Updating Method Based Android Malware Detection for IoT Services " 978-1-5386-7747-6/19/$31.00 ©2019 IEEE

[4]. Shengchu Zhao1, Wei Li1, Tanveer Zia2 and Albert Y. Zomaya1 "A Dimension Reduction Model and Classifier for Anomaly-Based Intrusion Detection in Internet of Things " 978-1-5386-1956-8/17 $31.00 © 2017 IEEE

[5]. Mohmudul hasan, Md. Milon, Md. Milon Islam, Md Ishrak Islam Zarif, M.M.A. Hashem "Attack and anomaly detection in IoT Sensors in IoT sites using machine learning approaches" elsevier ScienceDirect 2019

[6]. Prachi Shukla "ML-IDS: A Machine Learning Approach to Detect Wormhole Attacks in Internet of Thing" 978-1-5090-6435-9/17/$31.00 2017 IEEE

[7]. Jadel Alsamiri1, Khalid Alsubhi2 "Internet of Things Cyber Attacks Detection using Machine Learning" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 12, 2019

[8]. Giampaolo Casolla,Salvatore Cuomo,Vincenzo Schiano di Cola , and Francesco Piccialli "Exploring Unsupervised Learning Techniques for the Internet of Things " 1551-3203 © 2019 IEEE

[9]. YU-XIN MENG "The Practice on Using Machine Learning For Network Anomaly Intrusion Detection" 2011 IEEE

[10]. Chi Cheng, Wee Peng Tay and Guang-Bin Huang "Extreme Learning Machines for Intrusion Detection" - WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, 2012 - Brisbane, Australia

[11]. Naeem Seliya , Taghi M. Khoshgoftaar "Active Learning with Neural Networks for Intrusion Detection" IEEE IRI 2010, August 4-6, 2010, Las Vegas, Nevada, USA 978-1-4244-8099-9/10/$26.00 ©2010 IEEE

[12]. Kamarularifin Abd Jalill, Mohamad Noorman Masrek "Comparison of Machine Learning Algorithms Performance in Detecting Network Intrusion" 201O

International Conference on Networking and Information Technology 978-1-4244-7578-0/$26.00 © 2010 IEEE

[13]. Shingo Mabu, Member, IEEE, Ci Chen, Nannan Lu, Kaoru Shimada, and Kotaro Hirasawa, Member, IEEE "An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming" IEEE, JANUARY 2011

[14]. Liu Hui, CAO Yonghui "Research Intrusion Detection Techniques from the Perspective of Machine Learning" - 2010 Second International Conference on MultiMedia and Information Technology 978-0-7695-4008-5/10 $26.00 © 2010 IEEE

[15]. Jingbo Yuan , Haixiao Li, Shunli Ding , Limin Cao "Intrusion Detection Model based on Improved Support Vector Machine" Third International Symposium on Intelligent Information Technology and Security Informatics 978-0-7695-4020-7/10 $26.00 © 2010 IEEE

[16]. Maria Muntean, Honoriu Vălean, Liviu Miclea, Arpad Incze "A Novel Intrusion Detection Method Based on Support Vector Machines" IEEE 2010.

[17]. W. Yassin, Z. Muda, M.N. Sulaiman, N.I.Udzir, "Intrusion Detection based on K-Means Clustering and OneR Classification" IEEE 2011.

[18]. Mohammadreza Ektefa, Sara Memar, Fatimah Sidi, Lilly Suriani Affendey "Intrusion Detection Using Data Mining Techniques" IEEE 2010.

[19]. https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iot.php

[20]. Hanwen Wang, Biao Han, Jinshu Su," Biao Han, Jinshu S" 978-1-5386-9380-3/18/$31.00 ©2018 IEEE

[21]. Ibraheem Aljamal, Ali Tekeoglu Korkut Bekiroglu, Sangupta "Hybrid Intrusion Detection System Using Machine Learning Techniques in Cloud Computing Environment" 978-1-7281-0798-1/19/$31.00 ©2019 IEEE SERA 2019, May 29-31, 2019, Honolulu, Hawai