International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 9, pp: 507-513 www.ijaem.net ISSN: 2395-5252

# "A Machine Learning Based Classification Framework for Intrusion Detection System in Internet of Things"

Pragya, Mr. Nitesh Gupta, Mr. Anurag Srivastava
*MTech Scholar Department of CSE NIIST Bhopal*
*Associate Professor Department of CSE NIIST Bhopal*
*HOD Department of CSE NIIST Bhopal*

Date of Submission: 15-11-2020 Date of Acceptance: 28-11-2020

**ABSTRACT**—If we talk about in the current situation of world, Internet is the rapid growing technology.Internet of Things (IoT) combines hundreds millions of devices which are capable for the interaction between users and devices. Internet of Things (IoT) produces a very large amount of raw data in the form of log files. IoT infrastructure increased in all aspects, threats and attacks in these infrastructures are also growing proportionally. Some Researchers are found that the combination of machine learning technologies with an intrusion detection system is an correct way to resolve the problems of traditional IDSs have when we are used for IoT. This research involves the design of a novel intrusion detection system for IoT purpose.This paper proposed the classification framework for improving intrusion detection system. This framework uses the log probability concepts in naive Bayes machine learning algorithms. Classification framework used the BoT-IoT dataset for experimental purpose. The experiment result show that newnaive Bayes algorithms gives better accuracy in comparison to old naive Bayes and other classification algorithms.

**Keywords—** Internet of Things (IOT), Classification, Intrusion detection system (IDS),Naïve Bayes, DataSets, Bot-IoT, Pre-processing, Knowledge Discovery in Database (KDD), Data Mining, Machine Learning.

## I. INTRODUCTION

The "Internet of Tings" (IoT) describes many different systems and devices that are constantly connected to Internet, giving information from their sensors or interacting with their actuators. By 2020 it is estimated that there will be 4.5 billion IoT connecting with Internet. These devices have special features, such as a low computing capacity and the use specific lighter protocols. This makes IoT devices more efficient, smaller, and less energy consuming; however these low settings reduce their encryption capacity. These heterogeneous systems and networks offer new challenges in cyber security, such as new vulnerabilities and anomalies. attacks in recent years exploited these vulnerabilities by carrying out distributed denial of service attacks infecting IoT devices and attacking with as many as 400,000 simultaneously connected devices.

Security of devices and data is becoming a more important issue in now days. Thetechnique of improving network security is the use of Intrusion Detection Systems (IDS). IDS are one of the most productive techniques for detecting attacks within a network. This tool can detect network intrusions and network misuses by matching patterns of known attacks against ongoing network activity. With this purpose, our focus is to develop an IDS with machine learning models for the IoT. IDS use two different detection methods: signature-based detection and anomaly-based detection. Signature-based detection methods are effective in detecting well-known attacks by inspecting network traffic for specific patterns. Anomalybased detection systems identify attacks by monitoring the behavior of the entire system, objects, or traffic and comparing them with a predefined normal status. Machine learning techniques are used to improve performance of detection methods. These anomaly-based IDS have good results in qualifying frames that may be under attack, and they are effective even in detecting zero-day attacks. To build a machine learning classifier it is necessary to use a dataset.

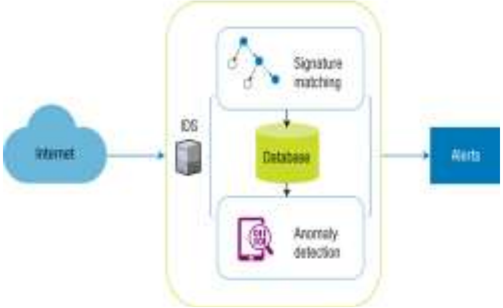

Figure : 1 IDS in IOT

### 1.1 IoT Components:

IoT components primarily include the following: Sensor- It is physical entity which senses the environment data, e.g. - temperature, air speed, humidity, movements. Actuator – it is responsible of movement in device when it get any control signal. For instance rotate the CCTV Camera in any direction. Network – IoT objects are tied up with networks by various wireless standards. 802.15 standard are using for wearable device, Zigbee or 802.11 used for home automation. Power efficient network standards have preferred mostly. User – people control the object via some user interface. User interface application provides facility to people to interact with devices. [2]

IoT devices cannot support complex security structures given their limited computation and power resources. Complex security structures of the IoT are due to not only limited computation, communication and power resources but also trustworthy interaction with a physical domain, particularly the behavior of a physical environment in unanticipated and unpredictable modes, because the IoT system is also part of a cyber-physical system; autonomously, IoT systems must constantly adapt and survive in a precise and predictable manner with safety as a key priority, particularly in settings where threatening conditions, such as in health systems, might occur. Moreover, new attack surfaces are introduced by the IoT environment. Such attack surfaces are caused by the interdependent and interconnected environments of the IoT. Consequently, the security is at higher risk in IoT systems than in other computing systems, and the traditional solution may be ineffective for such systems. IoT systems are accessible worldwide, consist mainly of Constrained resources and constructed by lossy links. Therefore, crucial modifications of existing security concepts for information and wireless networks should be implemented to provide effective IoT security methods.

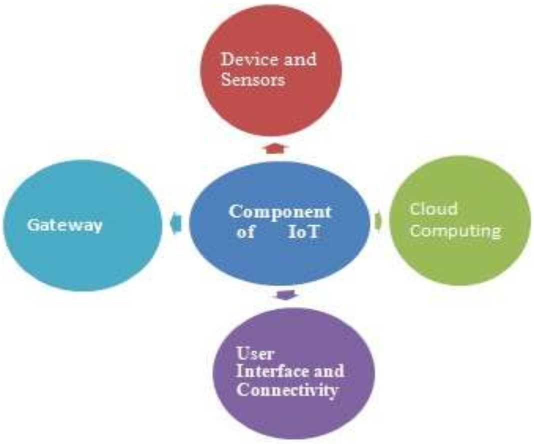

Figure :2 Component of IoT

The main focus of this paper is to develop a efficient framework for classification of IoT dataset and improve the accuracy of machine learning classifier in short time with low error rate. The structure of this paper is as follows: The Introduction outlines the context of this topic. next section discussed the previous work and concept of machine learning in brief are in section three. In next section proposed work, result and conclude the paper.

## II. LITERATURE REVIEW

In view of the fact that IoT represents a new concept for the Internet and smart data, itis a challenging area in thefield of computer science. The importantchallenges for researchers with respect to IoT consist of preparing andprocessing data and discovering knowledge.

In this research paper [2] the authors have used machine learning techniques, approaches or methods for securing things in IOT environment. This paper attempts to review the related research on machine learning approaches to secure IOT devices.

In this research [4] the various machine learningmethods that deal with the challenges presented by IOT data by considering smart cities as the main use case. Thekey contribution of this study is the presentation of taxonomy of machine learning algorithms explaining howdifferent techniques are applied to the data in order to extract higher level information. The potential andchallenges of machine learning for IOT data analytics will also be discussed. A use case of applying a SupportVector Machine (SVM) to Aarhus smart city traffic data is presented for a more detailed exploration.

In this research paper [5] authorsaims to provide a brief overview of machine learning methods for internet of things (IOT). Authors present some of the applications of machine learning in IOT and have tried to provide an overview of the types of ML, ML task and its applications as related to IoT. In conclusion, it is needful to mention that ML provides higher precision in calculations and for prediction, it is highly effective and is able to look at a lot of information in smaller interims of time.

In the research paper [8] authors review ML/DLmethods for IoT security and present the opportunities,advantages and shortcomings of each method. Authors discuss theopportunities and challenges involved in applying ML/DL to IoTsecurity. These opportunities and challenges can serve as potentialfuture research directions.

This research paper [9] addresses the comparison of severalfrequently used ML classifiers from the group of SVMlike classifiers, namely SMO and C-SMV algorithm, and arange of ensemble algorithms on the other side, namelyLAD Tree, REPTree, RF and MultiBoost. The analysis isbased on a range of testing procedures in Weka, with agoal to estimate a set of selected performance metrics andmake classifier comparison. As the analysed UNSWNB15dataset belongs to a unbalanced dataset category,for the proper examination of the classifiers we haveassumed the need for calculating the precision, recall,ROC and necessary time for classification.

## III. INTRUSION DETECTION APPROACH FOR IOT

The proposed intrusion detection system (IDS) essentially targets smart places connected to IoT devices. Its main goal is to detect potential attacks that can occur through wireless communications.intrusion detection system that can detectcomplex 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. The core network of the Internet of things is still a traditional network but it has more complexities. The large number of nodes in the Internet of things makes the network more vulnerable, and the impact of attacks can be more serious than for conventional networks. The performance of traditional intrusion detection methods will be greatly reduced in this complex environment [1,14]. At present, intelligent, distributed intrusion detection has become a hot topic.

## IV. 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. 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) 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.

### 4.1. Types of Machine Learning

ML mainlydivided into three categories. Supervised and unsupervised arewidely used categories. In supervised machine algorithm, training data has input and its corresponding output.Unsupervised machine learning, we do not have any output.In reinforcement machine learning a software agentautomatic take action to maximize the performance or award. For active learning type, a PC can simply get information for a confined game plan of cases. Exactly when used instinctively, this information can be shown to the customer.

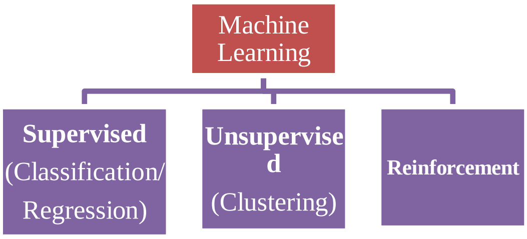
Machine Learning
Supervised (Classification/ Regression)
Unsupervise d (Clustering)
Reinforcement

Figure 3: Types of Machine Learning

- **Supervised learning :**In this type of learning, theoutput class labels of the data are known or can be calculated. In cases where the labels are unknown,their operational data will be available.
- **Unsupervised learning:** No imprints, labelling orcategorization are given to the learning computation.It isolates the information to find the structure in itsdata.
- **Reinforcement learning**: It is an area of Machine Learning. It is about taking suitable action to maximize reward in a particular situation.

### 4.2 Real Machine learning workflow

- Gathering Data
- Cleaning data
- Model building and choosing the correct calculation
- Gaining insights from the outcomes
- Visualizing the information

### 4.3 Machine Learning Algorithms:-

### A. Naive Bayes

Naive Bayes classifier infers that for a given class,features are independent [12]. Using the most frequent values of the features naive Bayes classifier dispense theclass label to the instances [13]. It calculates the priorprobability of each class in the training phase using theoccurrences of the each feature for each class. NaiveBayes finds the posterior probability of the class based onthe class prior probability [14]. It deduce that the result ofthe predictor for a given class is independent of the valuesof other predictor. Using the aforementioned probabilitiesit assigns the class label to the new data.

### B.Support Vector Machines (SVM)

SVM is a supervised ML algorithm with low computational complexity, used for classification and regression. It has the ability to work withbinary as well as with multi-class environments. It classifiesinput data into n dimensional space and draws n − 1 hyperplane todivide the entire data points into groups.

### C. J.48

**J**.48 is a type of decision tree. Decision tree considers theclass as a dependent variable which lies on the leaf of atree. Decision tree is a graphical representation of theclassification algorithm [15]. J.48 creates, first, a decisiontree in order to classify new instances. Dependentvariables (classes) are decided by the values of theinternal nodes which represent the variables which areconsidered independent variable.

### 4. The BoT-IoT Dataset

Many researchers uses the BoT-IoT dataset for their research in IDS for IoT. 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 argues 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 exfiltration 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. In this work used csv data file with all features and about five lakhs instances contain 46 features.

### A. Log Probability

A log probability is simply the logarithm of a probability. The use of log probabilities means representing probabilities in logarithmic space, instead of the standard [0, 1] interval. In most machine learning tasks we actually formulate some probability p which should be maximized, here we would optimize the log probability log(p) instead of the probability for class θ. The use of log probabilities determines better numerical stability, when the probabilities are close to each other and very small.

**ex = y**

**loge (y) = x**

Where x is probability. To get back the values of probability take log of y on base e.

### 4. PROPOSED WORK

Some of the researchers in the field of machine learning has addressed the strategy for improve the performance of ML classifier which is used in modern intrusion detection system. To classify abnormal behaviorand minimizing misclassification propose a classification framework based on new naïve Bayes algorithm are proposed. The Proposed naïve Bayes algorithm is used the concept of log probability. Detail about the log probability discuss in Introduction.

```
Proposed new naïve Bayes Algorithm:
Old Naive Bayes Algorithm
Begin To get Class of specific Instance
state Probabilities of Array size = n
(n = total  number of classes in dataset)
Loop For j=0 to n-1
For each class get value of probability and save in probability [j]
End For
Get no. of attributes
Loop, While
Declare variable temp and max=0;
Loop For j=0 to n-1
Get probability estimates of each attribute and product over of these with each class probabilities.
Get max of these probability obtained in previous step and store in array of probabilities.
Now get / Take log of probabilities and update in array of probabilities.
Take max value from array of log of probabilities
End For
End while
```

This is proposed new naïve Bayes algorithm which is used for improving the IDS performance. We used BoT IoT dataset. The first step is pre-processing in this step clean the raw data and get ready to processed now in attribute extraction steps select appropriate attribute from dataset. In the next step, we applied new naive Bayes classification algorithm on training and testing dataset in order to classify normal and abnormal data and measure performance. This same process also applied for general naïve Bayes classifier algorithms and compare result. Architecture of the proposed work are shown in figure 2. For experiment purpose weka 3.8 tool is used.

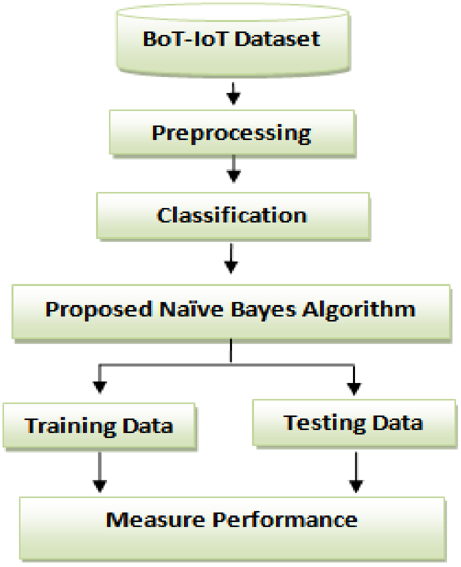

**Figure 4: Proposed Classification framework**

## V. RESULT ANALYSIS:

The experimental result of the proposed Intrusion detection model for Internet of things show that with proposed model J48 classifier gives better accuracy and take very less time to build model and improve the IDS performance. Experimental result also compare with the performance of new naive Bayes and general naive Bayes algorithms. The performance parameter are as follow as: accuracy, error rate and time taken to build model. Table 1 show the comparison of experimental result. Result show that in comparison to general naive Bayes multinomial text proposed new naive Bayes algorithm give better accuracy and less error rate. time take to build model for new naive Bayes is little bit max to general naive Bayes.

**Table 1: Comparison of result**

| **Parameter** | **New naïve Bayes** | **General naïve Bayes multinominaltext** |
|---|---|---|
| **Accuracy** | 82.86 % | 79.25 % |
| **Error Rate** | 17.13 % | 20.74 % |
| **Time taken to build model** | 13.33 Second | 0.14 Second |

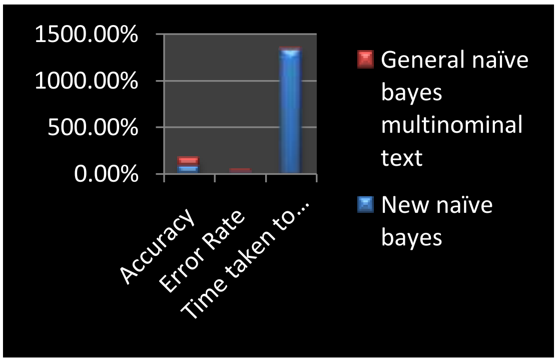

**Figure 5: Shows Result Comparison Graph**

## VI. CONCLUSION

Machine learning techniques are used for classification of data. Many existing study about the IDS are show that machine learning algorithms are used for classification of normal and abnormal data from large dataset. In this work new naïve bayes classification algorithms based on log probability is proposed. With the help of this naïve bayes classifier, IDS improve the performance. Proposed work improves the performance of classifier which classifies the abnormal association, high accuracy and detection rate with low false alarm. The proposed work is completed by telling a framework for Classification and method to evaluate the framework. The issue of correct classification and model building time is also important for evaluating the framework. Proposed framework with new naïve bayes classification algorithms is showing greater accuracy when tested with general naïve bayes classifiers.

## REFERENCES

[1]. D. Evans, "The Internet of Things How the Next Evolution of the Internet is Changing Everything," CISCO, 2011.

[2]. AmitSagu, Nasib Singh Gill "Securing IoT Environment using Machine Learning Techniques" International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-3, February, 2020

[3]. 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

[4]. YueXu "Recent Machine Learning Applications to Internet of Things (IoT)" Recent Machine Learning Applications to Internet of Things (IoT) Recent Machine Learning Applications to Internet of Things (IoT)

[5]. Mohammad SaeidMahdavinejadMohammadrezaRezvan MohammadaminBarekatainPeymanAdibiPayamBarnaghiAmit P. Sheth[1,2][3][4] "Machine learning for internet of things data analysis: a survey" http://www.keaipublishing.com/en/journals/digital-communications-and-networks/

[6]. Arun Kumar Rana1, AyodejiOlalekan Salau2, Swati Gupta3, Sandeep Arora4 "A Survey of Machine Learning Methods for IoT and their Future Applications"

[7]. Amity Journal of Computational Sciences (AJCS) Volume 2 Issue 2 ISSN: 2456-6616 (Online)

[8]. 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

[9]. Mohammed Ali Al-Garadi, Amr Mohamed, Abdulla Al-Ali, Xiaojiang Du, Mohsen Guizani "A Survey of machine and deep learning methods for internet of things (IoT) Security"

[10]. ValentinaTimcenko, SlavkoGajin "Machine learning based network anomaly detection for IoT environments" ieee explorer

[11]. 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

[12]. GiampaoloCasolla,SalvatoreCuomo,VincenzoSchiano di Cola , and Francesco Piccialli "Exploring Unsupervised Learning Techniques for the Internet of Things " 1551-3203 © 2019 IEEE

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

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

[15]. 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

[16]. 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

[17]. 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

[18]. 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

[19]. 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

[20]. Maria Muntean, HonoriuVălean, LiviuMiclea, Arpad Incze "A Novel Intrusion Detection Method Based on Support Vector Machines" IEEE 2010.

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

[22]. MohammadrezaEktefa, Sara Memar, Fatimah Sidi, Lilly SurianiAffendey "Intrusion Detection Using Data Mining Techniques" IEEE 2010.

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

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

[25]. IbraheemAljamal, 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

[26]. Fatima Hussain, Rasheed Hussain, Syed Ali Hassan, and EkramHossain "Machine Learning in IoT Security: Current Solutions and Future Challenges " arXiv: 1904. 05735v1 [cs.CR] 14 Mar 2019

[27]. NickolaosKoroniotis, NourMoustafa, Elena Sitnikova, BEnjamin Turnbull "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset" https://doi.org/10.1016/j.future.2019.05.041 0167-739X/© 2019 Elsevier B.V. All rights reserved.