National Conference on Innovations in Computing and Networks (NCICN) - 2019
© 2019 IJSRSET | Volume 5 | Issue 9 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099

A Two Stage Method for Classifying the Cyber Attacks in Amazon Web Service Cloud By Machine Learning Techniques

Yasir A*[1], Nishanth M Raj*[2], Thasnim L*[2],Thouheed B*[2]

*[1]Asst. Professor, Department of Computer Science and Engineering, Younus College of Engineering and Technology, Kollam, India

*[2]B.Tech, Department of Computer Science and Engineering, Younus College of Engineering and Technology, kollam, Kerala, India

ABSTRACT

In Modern digital world, security of our valuable information is always an essential issue. There may chances of deferent cyber attacks. Against these, IDS and many security techniques have been used. To obtain high detection rate and low false alarm rate the researchers also been used data mining techniques and other Machine Learning (ML) techniques. Machine learning is a type of artificial intelligence(AI) that provides computers with the ability to learn without being explicitly programmed.This paper proposed to classify the cyber attack in Amazon Web Service Cloud by using Multiclass SVM and Pattern Matching ML techniques.The system uses a two stage method for classification. The first stage classify the maximum number of attacks by using Multiclass SVM algorithm. The second stage is Pattern Matching to classify and prevent the remaining attacks. The existing ML techniques do not provide well processing of large data sets because of the network traffic. The Proposed method is an cloud based ML technique, which uses NSL KDD CUP99 dataset.The model is evaluated in terms of accuracy with the benchmark.

Keywords: Amazon web service , Machine learning, Multiclass support vector machine(MCSVM), Pattern Matching Algorithms.

I. INTRODUCTION

Security of valuable information is always a very essential issue for modern digital world. Intrusion Detection System (IDS) and many security techniques is widely used against cyber attacks. In recent years, the number of attacks on the computer networks and its components are getting increasing. To protect from these attacks various Intrusion detection techniques have been used. Intrusion Detection System (IDS) is a system which collects and analyzes the information from the network to identify various attacks made against the components of a network. There are many types of Cyber Attacks like Brute Force, Probing, Denial of Service, Bot

Attack, Phishing, SQL Injection. Cyber attack detection system inspects all inside and outside network movement and identifies mistrustful patterns that may point to a network or system attack from someone attempting to break into or compromise a network. In cyber attack detection, we generally deal with a large amount of data collected from cyber agent to make a decision on the current situation of the network. Different types of attacks may have different effects

on the operations of a cyber network. As a result, the data that need to be collected from cyber agent vary from one kind of attacks to the other.

Cloud is a Secure Layer.Cloud computing is a computing paradigm, where a large pool of systems are connected in private or public networks, to provide dynamically scalable infrastructure for application data and file storage.
Cloud Providers offer services that can be grouped into three categories.

- ✓ Software as a Service (SaaS): In this model, a complete application is offered to the customer, as a service on demand.
- ✓ Platform as a Service (Paas): Here, a layer of software, or development environment is encapsulated offered as a service, upon which other higher levels of service can be built.
- ✓ Infrastructure as a Service (Iaas): IaaS provides basic storage and computing capabilities as standardized services over the network.

The customer would typically deploy his own software on the infrastructure. Some common examples of cloud are Amazon, GoGrid, 3 Tera, etc. Enterprises can choose to deploy applications on

Public, Private or Hybrid clouds. Cloud Integrators can play a vital part in determining the right cloud path for each organization. Public clouds are owned and operated by third parties; they deliver superior economies of scale to customers, as the infrastructure costs are spread among a mix of users, giving each individual client an attractive low-cost, "Pay-as-you-go" model. There are two variations to a private cloud.

- ✓ On-premise Private Cloud: On-premise private clouds, also known as internal clouds are hosted within ones own data center.
- ✓ Externally hosted Private Cloud: This type of private cloud is hosted externally with a cloud provider, where the provider facilitates an exclusive cloud environment with full guarantee of privacy.

AWS [Amazone web service cloud] provides you with guidance and expertise through online resources, personnel, and partners. AWS provides you with advisories for current issues, plus you have the opportunity to work with AWS when you encounter security issues. You get access to hundreds of tools and features to help you to meet your security objectives. network security, configuration management, access control, and data encryption. Finally, AWS environments are continuously audited, with certifications from accreditation bodies across geographies and verticals. In the AWS environment, you can take advantage of automated tools for asset inventory and privileged access reporting.

- ✓ Keep Your Data Safe: The AWS infrastructure puts strong safeguards in place to help protect your privacy. All data is stored in highly secure AWS data centers.
- ✓ Meet Compliance Requirements: AWS manages dozens of compliance programs in its infrastructure.
- ✓ Save Money: Cut costs by using AWS data centers. Maintain the highest standard of security without having to manage your own facility
- ✓ Scale Quickly: Security scales with your AWS Cloud usage. No matter the size of your business, the AWS infrastructure is designed to keep your data safe.

II. RELATED WORKS

Network security starts with authorization, commonly with a username and a password. Network security consists of the provisions and policies adopted by a network administrator to prevent and monitor unauthorized access, modification in system, misuse, or denial of a

computer network and network-accessible resources. Traditional Intrusion detection system using data mining and machine learning techniques are work on information system they are not working on cloud environment [5]. Here give some literature about Intrusion detection system and using cloud for classification with machine learning algorithms. 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[10].The data flows through a network is very large and it is difficult to analyze and detect the attacks using traditional methods. Multiple choices of cloud computing models are performance and computational requirements. The popular statistical tools and environments like Octave, R and Python are now embedded in the cloud as well [1]. Anku Jaiswal [2] proposed the cloud based attack system. Authors add new valued feature to the cloud- based websites and at the same time introduces new threats for such services. DDoS attack is one such serious threat. Covariance matrix approach is used in

III. MATERIALS AND METHOD

Dataset

The Proposed system uses NSL KDD CUP99 dataset.In Earlier days the researcher focused on DARPA dataset for analyzing intrusion detection. It consists of seven weeks of training and also two weeks of testing raw tcpdump data. The main drawback is its packet loss. To solve these issues, a new data set as, NSL-KDD is proposed, which consists of selected records of the complete KDD data set. It have 5 million single connections for training and 2 million connections for dataset.testing. It affects the Performance of the system.So we use improved version of KDD. It can mostly used data set for evaluation of anomaly detection methods. The advantages of this dataset is,no redundant copies in training set and hence no duplicate copy in testing,will not produce any biased result.And also it have better reduction rates.The number of selected records from each difficult level group is inversely proportional to the percentage of records in the original KDD data set.

Multiclass SVM

Support Vector Machine is a machine learning approach used for classification and regression analysis. It depends on supervised learning models and trained by learning algorithms. They analyze the large amount of data to identify patterns from them.An SVM generates parallel partitions by generating two parallel lines. For each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible.Do this partitioning by a plane called hyperplane.Support vector machines (SVM) were originally designed for binary classification. Several methods have been proposed where typically we construct a multi-class classifier by combining several binary classifiers.Here, we compare their performance with three methods based on binary classifications: “one-against-all,” “one-against-one,” and DAGSVM. Our experiments indicate that the “one-against-many” and DAG methods are more suitable for practical use than the other methods. Results also show that for large problems methods by considering all data at once in general need fewer support vectors. At a time,chances of many attack are there,this help to classify each attack at a time.This algorithm helps to train the classifier and classify the attack.At a time, one trains K (K – 1)/2 binary classifiers for a K-way multiclass problem; each receives the samples of a pair of classes from

the original training set, and must learn to distinguish these two classes. At prediction time, a voting scheme is applied: all K (K – 1) / 2 classifiers are applied. The figure the comparison of accuracy of SVM with other networks.

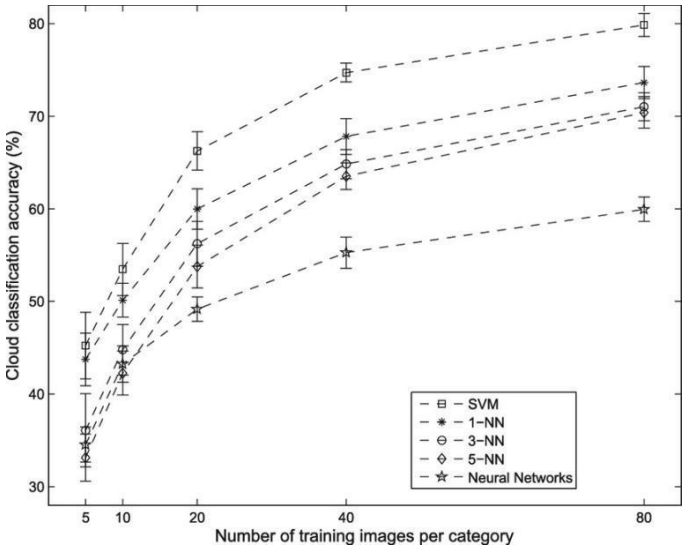

Fig 2. comparison chart

Pattern Matching Algorithm

Pattern Matching algorithm is used to develop password equivalent to a corresponding pattern. When a user login into a application or site thus we convert the hash code of password entered. It Identifies the pattern is matched with the login process and check whether the pattern is matched. Then equivalent password is generated,by depending the pattern matching process.It automatically generate the chances of dummy password,depends on the hash code of real password.When an attacker try to login with multiple password,the system automatically considered as a malicious activity. Then the system block the corresponding IP and potentially suspend the account.For a period,the account is blocked even real user login.

IV. METHODOLOGY

The machine learning process, which includes;

- Input Stage
- Preprocessing Stage
- Apply Classification Algorithms

The machine learning process in this system contains defferent stages.Intially,input the dataset or the input data,then preprocess this data.Data preprocessing include data cleaning,data integrity,data redundancy.There is two stages of machine learning methods,which is MCSVM and Pattern Matching algorithm.First stage classify the maximum number of attacks by using MCSVM and the attack that cannot be detected by MCSVM is detected and prevented by using Pattern Matching algorithm.Here,apply Machine Learning Algorithm and an ensemble method for better accuracy in classifying and detecting the attack. The execution flow of the system,

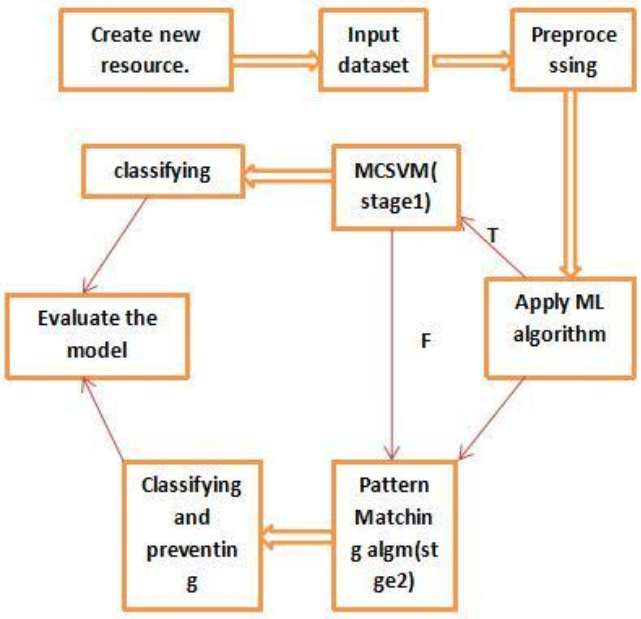

Figure 1: Execution flow

Algorithm

It implies execution of Implemented Work (Experiment Steps):The experimental steps that are represented and explained below:

Step 1: Create New Resource
Step 2: Import/Upload the dataset. Step 3: Pre-process the dataset
Step 4: Randomly split and partition the data into 70%
training and 30% testing, using the 'Split Data' module. Step 5: Identify categorical attributes and cast them into
categorical features using the 'Edit Metadata' module.
Step 6: Convert to Indicator Values module to convert columns.
Step 7: Select Columns in Dataset those are relevant
Step 8: Apply Ensemble Method
Step 9: Apply Machine Learning Algorithm to Train the model.
Step 10: Now Score and Evaluate the Model.
Step 11: The 'Evaluate model' also visualizes the results through confusion matrix.

MCSVM helps to classify the attack and Pattern matching is used to develop equivalent pattern to the corresponding pattern. The dummy passwords are created instead of real passwords .Any hackers tries to login an account ,it considered as a malicious activity.Create the new resource for upload the dataset and pre-process it.The dataset is split for training and testing the data.The categorical attributes are identified and cast them to convert into Indicator value.The converted coloumns are selected for applying machine learning technique and ensemble methods for better accuracy.Finally evaluate the model.

PERFORMANCE METRICS

The metrics used for the research work is described in this section.

Precision

Precision is the part of significant instances between the retrieved instances. The Eq.of precision is,

Precision = TP/(TP+FP)

Recall

Recall is the small part of appropriate instances that have been retrieved over the total quantity of relevant instances.

The Eq.of recall is,

Recall = TP/(TP + FN)

Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same.

Accuracy = TP+TN/TP+FP+FN+TN The work proposes following types of attacks:

- Dos attack: DoS attack is a type of attack in which the hacker makes a computing or memory resources too busy or too full to serve legitimate networking requests and hence denying users access to a machine.
- Bot attack
- Phishing attack
- Probing attack: Scan networks to gather deeper information.

- U2R attack: Illegal access to gain super user privileges.
- R2L attack: It has been widely known to be launched by an attacker to gain unauthorized access to a victim machine in the entire network.
- Sql injection: It is the vulnerability that results when you give an attacker the ability to influence the Structured Query Language (SQL)queries that an application passes to a back-end database.

V. RESULT AND DISCUSSION

The experiment is evaluated on a simple multiclass SVM algorithm by accuracy parameter. Accuracy is well more for this proposed method when compared with existing method. The results obtained using the benchmark code by setting the support vector machine model. The results obtained using the benchmark code or traditional method got the accuracy of 0.50241 in experiment in detecting the cyber attack, while the proposed system uses two stages of algorithms which helps in detecting and preventing the cyber attack, results with 0.9633 accuracy.MCSVM helps in classifying and detecting the attack at an accuracy of

0.8409 (84%) and Pattern Matching algorithm gives better accuracy of 0.9633.By considering both the accuracy level,the system achieves 0.96332 of accuracy. The IP of unauthorised user is blocked by Pattern Matching Algorithm and is potentially suspended for a period. The system is implemented in cloud which gives more secure for our valuable information.

Overall accuracy (Proposed system)	0.9633
Accuracy(Existing system)	0.50241
Recall	0.963326
Precision	0.9626

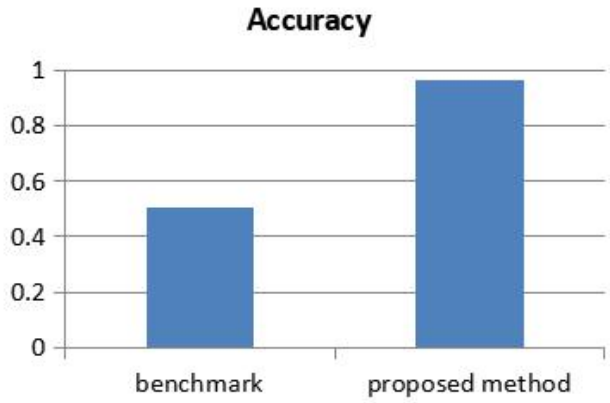

Figure 2:Comparison chart

VI. CONCLUSION AND FUTURE WORK

The paper proposes a new Cloud Based Machine Learning technique for improving the Cyber Attack classification. The model classify the attacks using "multiclass SVM" and Pattern Matching algorithm with high accuracy than benchmark. The benchmark with the accuracy of 0.50241 and our method have the accuracy of 0.9633.The model have low false alarm rate and high detection rate for many attack classification. In future work our system aims to design a model that can be optimized to handle attack classification in vaious platforms and domains. Also, the model can be modified for applying on Hadoop MapReduce platform for more accuracy.

VII. ACKNOWLEDGMENT

We are sincerely thankful to our Principal, Dr. P Sreeraj, for providing us the facilities in order to go ahead in the development of our research. We express our deep and sincere gratitude to Dr. Nijil Raj N, Head of Department, Computer Science and Engineering and Prof.Yasir A,Guide and Project Coordinator,Computer

Science and Engineering for providing valuable advice and timely instructions.We would like to express our very great appreciation to Mr.Sreedharsh S for his valuable and constructive suggestions during the planning and development of this

research work. His willingness to give his time so generously has been very much appreciated.

VIII. REFERENCES

[1]. Anku Jaiswal, Chidananda Murthy P, Madhu BR "Prevent DDOS Attack in Cloud Using Machine Learning" Volume 6, Issue 6, June 2016 ISSN: 2277 128X

[2]. https://portal.azure.com.

[3]. Ms. Parag K. Shelke, Ms. Sneha Sontakke, Dr. A. D. Gawande "Intrusion Detection System for Cloud Computing" International Journal of Scientific & Technology Research Volume 1, Issue 4, May 2012 ISSN 2277-8616

[4]. Amjad Hussain Bhat1, Sabyasachi Patra2, Dr. Debasish Jena3 "Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines" Web Site: www.ijaiem.org Email: editor@ijaiem.org, editorijaiem@gmail.com Volume 2, Issue 6, June 2013. [5] Solane Duquea*, Dr.Mohd. Nizam bin Omarb Using Data Mining Algorithms for Developing a Model for Intrusion Detection System (IDS) www.sciencedirect.com, Elsevier 2015.

[5]. Kamarularifin Abd Jalill, Mohamad Noorman Masrek "Comparison of Machine Learning Algorithms Performance in Detecting Network Intrusion" 201O International Conference 0n Networking and Information Technology 978-1-4244-7578-0/$26.00 © 2010 IEEE

[6]. Daniel Pop, "Machine Learning and Cloud Computing Survey of Distributed and SaaS Solutions", https://www.researchgate.net/publication/257068169.

[7]. 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

[8]. Shane Miller, Kevin Curran, Tom Lunney " Cloud- based machine learning for the detection of anonymous web proxies" ISSC 2016.

[9]. Ch.Ambedkar, V. Kishore Babu, "©ARC Page 25 Detection of Probe Attacks Using Machine Learning Techniques" International Journal of Research Studies in Computer Science and Engineering (IJRSCSE) Volume 2, Issue 3, March 2015, PP 25-29 ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)

[10].