Machine Learning Techniques For Cyber -Attacks Detection

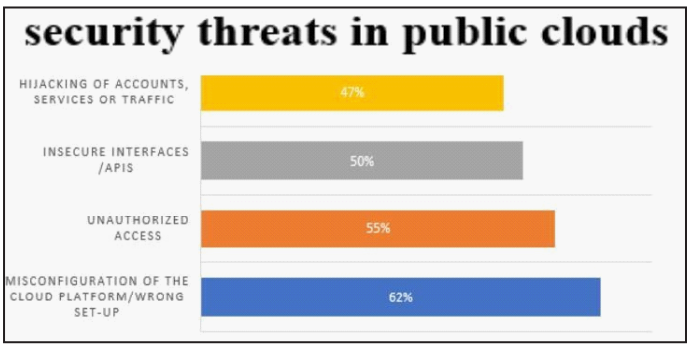
[1]GudlaAasrita, [2]D S Ramkiran
[1,2]Dept. of CSE, KIET, Kakinada, AP, India

Abstract
Cloud computing is an evolving technology thatprovides reliable and scalable on-demand resources and different services to users with less infrastructure cost. Even though the cloud has many advantages it faces many drawbacks like vulnerability to attacks, network connectivity dependency, downtime, vendor lock-in, limited control. From the above-mentioned disadvantages, a security attack is the main drawback in the cloud. There are various security attacks like Denial-of-service (DOS) attack, Malware injection attack, Side channel attack, Man-in-the-middle attack, Authentication attack. To detect this attack in the cloud the machine learning algorithm like Support vector machine (SVM), Naive Bayes, Decision tree, Logistic regression, Ensemble methods can be used. In this paper, we have mainly focused on various security attacks in the cloud and the machine learning algorithms used for detecting the attacks.

Keywords:
Security attacks, Machine learning algorithms, Detection.

I. Introduction

The cloud is a booming technology in the computer sector. It refers to the accessing of the information technology and the software applications through the internet connection. The Software as a service (SAAS), Platform as a service (PAAS) and the Infrastructure as a service (IAAS) all together encapsulate to form the cloud. All the above services are the three types of services that is been provided by cloud computing. The services are hosted at the data centre by the cloud service providers for the organization or the individual users to utilize the services through a network connection. The cloud service providers are the companies that offer different services in the cloud. The major cloud service providers include AWS, Sales force, Cisco, Apple, Google, IBM (Soft Layer), Oracle, Microsoft (Azure), and SAP, Rack space, and Verizon (which acquired Terre mark. But the Sales force and the Apple are interested in providing their own application rather than hosting applications for others. The companies like Google, IBM, Microsoft, SAP provides all the three services of the cloud while the other companies provide either two or one of the cloud services. One of the disadvantages in cloud computing is security attacks. This drawback is due to the data storage at different geographical areas in cloud computing.



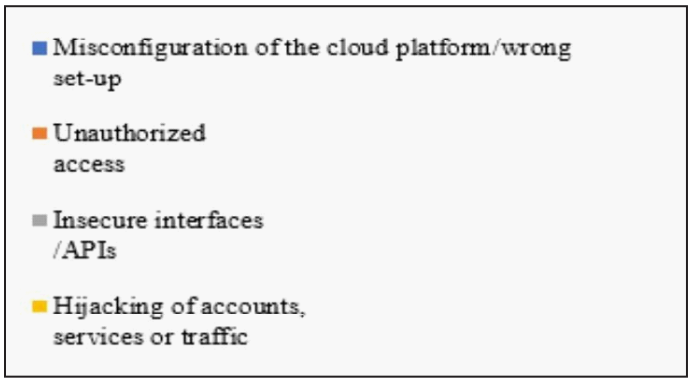


Fig. 1

The above chart describes the various security threats in public clouds as per the cloud security report provide by cloud security insiders, thus from the chart the misconfiguration of the cloud platform is about 62%, unauthorized access is about 55%, Insecure interfaces /APIs is about 50%, Hijacking of accounts, services or traffic is about 47%.

In section 2, we discussed different types of attacks on the cloud such as denial of service, malware injection attack, side channel attack, authentication attack, a man in the middle attack. Section 3, describes various machine learning algorithm used in security attack to detect like naive Bayes, support vector machine (SVM), K-means clustering, fuzzy logic, decision tree, and genetic algorithm.

II. Attacks On Cloud

The cloud encounters many security attacks due to its disadvantages. The various cloud attacks like Denial of service attack, Malware injection attack, side channel attack, Man in the middle attack and the authentication attack are discussed below. The attacks may happen at different parts of the cloud like the data storage, during a transaction, during resource utilization and sharing. The loss of the attack can be lower to higher based on the type of attack. The reason for the attack in the cloud is due to the huge increase in the use of cloud services.

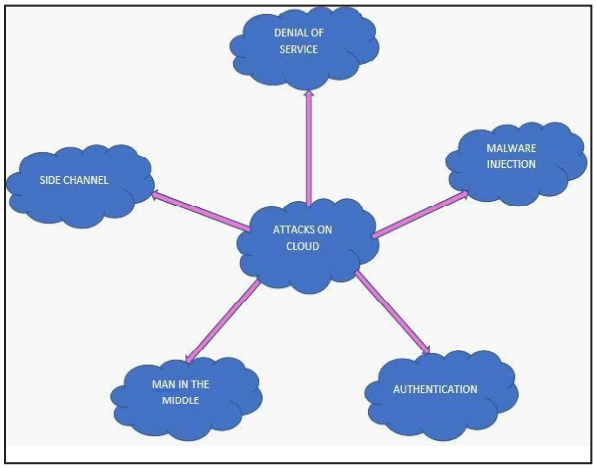


Fig. 2 Attacks on cloud

A. Denial of Service Attack

Denial of service attack the targeted cloud system is overloaded with the service requests from the attacker that stops it from responding to the upcoming new requests and to its users. According to some of the cloud security alliance, this cloud is very much vulnerable to this Dos attack. The Denial of service attack can be categorized into the DoS attack and the DDoS (Distributed denial of service attack). The attack was done using the single system and the single network is known as the DoS attack. The attack was done using multiple systems and the multiple networks are known as the Distributed denial of service attack (DDoS). The different types of the DDoS attacks are Volume based attack, protocol attacks, Application layer attack.

B. Malware injection attack

Malware injection attack the attacker injects the victim system with the malicious service or the malicious virtual machine. Here the attacker creates its own malicious virtual machine or the malicious service module and tries to add it into the cloud system. Then the attacker must behave so as to make the cloud system believe that it is a valid service. If the attacker succeeds then the cloud automatically redirects all the requests to this malicious service. Now the attacker can access the service requests of the victim services.

C. Side channel attack

The attacker tries to compromise the cloud system by placing a malicious virtual machine nearby to the target cloud system then it dispatches the side channel attack. These channels are created in the software implementation of cryptographic algorithms. Its impact may be greater than any other attacks as they attempt to retrieve secret data without any special privileged access and in a non-exhaustive manner. There are different categories of side channel attack like Timing attacks, Cache attacks, Electromagnetic attacks, and Power-monitoring attacks. Electromagnetic

attacks and power – monitoring attacks are mostly applicable to physical devices such as smart cards. The cache attacks and the Timing attacks are mainly applicable to the cloud computing.

D. Authentication attack

The Authentication attack mainly focuses on the authentication part of the cloud services. The primary authentication in most of the services is the username and the password which is a type of the knowledge-based authentication. The secondary authentication like shared secret questions, site keys, virtual keyboards is used by secure functioning organizations like the financial company. Some of the authentication attacks are the Brute Force Attacks, Dictionary Attack, Shoulder Surfing, Replay Attacks, Phishing Attacks, Key Loggers.

1. Brute force attack: This attack is like a trial and error method; all possible combinations of the password are applied to break the password.
2. Keyloggers: It is a form of a software program, where it monitors the actions of the user by recording each and every key pressed by the user.
3. Phishing attack: In this attack, the attacker redirects the user to the fake websites to get the passwords and the pin codes of the user, it is a kind of the web-based attack.

E. Man-in-the-middle-attack

Man-in-the-middle attack the attacker intercepts the message in the public key exchange and retransmits it by substituting its own public key for the requested one, but the two original are still communicating normally. The sender does not know that the messages sent by him is received by an attacker and he can access data, modify the message before retransmitting it to the receiver. Some of the man-in-the-middle attacks are Address Resolution Protocol Communication (ARP), ARP Cache Poisoning, DNS Spoofing, Session Hijacking.

III. A Machine Learning Algorithm For Detection

The machine learning algorithm allows software applications to produce accurate predicting outcomes without being explicitly programmed. The machine learning algorithm can be divided into classification algorithms and clustering algorithms. Some of the classification algorithms are the Naïve Bayes, support vector machine (SVM), decision tree, logistic regression, and ensemble methods. In this paper, we are going to use the classification algorithm.

A. Naïve Bayes

Naive Bayes depends on the Bayesian technique for playing out the classification process. It is a basic and most straight forward procedures for building classifiers models that allocate class labels to issue instance, represented as vectors of highlight values when the class labels are drawn from some limited set. The use of hidden Naive Bayes (HNB) gives exact outcomes than the traditional naive Bayes model. HNB can be connected to intrusion detection issues that experience the ill effects of dimensionality exceptionally related highlights and high system information stream volumes. Dos attack is distinguished utilizing 3 system: Multilayer perceptron (MLP), Naive Bayes and Random forest. MLP demonstrated the most elevated exactness rate 98.63% when contrasted with different systems. Display utilized naive Bayes classifier with k2 learning process on decreased NSL KDD dataset for each attack class. In the proposed model each layer is prepared to dataset a solitary sort of attack. The result of one layer is passed on to another layer to build the identification rate. It distinguishes attack that happens in an unverifiable circumstance.

B. Support Vector Machine (SVM)

SVM is used in classification and regression. classification can be viewed as the task of separating classes in feature space. It became famous when using the image as input, it gave good accuracy. Currently, SVM used in object detection and recognition, content-based image retrieval, text recognition, biometrics, speech recognition etc. Svm is a practical learning method based on statistical learning theory. Construct a hyperplane in the decision surface in such a way that the margin of separation between positive and negative. The goal of SVM is to find the particular hyperplane of which the margin is maximized. The particular data point for which the first or second line of the equation is satisfied with the equality sign is called a support vector.

C. Decision Tree

The decision tree algorithm is a kind of the classification-based machine learning algorithm. A decision tree is a flow-chartlike hierarchical tree structure which is composed of three basic elements: decision nodes corresponding to attributes, edges or branches which correspond to the different possible attribute values. The third component is leaves including objects that typically belong to the same class or that are very similar. Tree induction algorithms like Id3 and C4,5 create decision trees, it

takes only one attribute at a time. The decision tree nodes are created by choosing an attribute from the feature space of the dataset that brings maximum information gain by splitting the data on its distinct value. After the split, the information gain is calculated as the difference between the entropy of the initial dataset and the sum of the entropies of each of the subsets.

D. Logistic Regression

The logistic regression is the commonly used tool for discrete data analysis. It uses an equation as the representation. Logistic regression is used for predicting the probabilities of the various classes does an analysis and give a group of independent variables. It makes use of a linear equation with independent predictors for predicting a value. The predicted value can be anywhere from negative infinity to positive infinity of the system. We can squash the output of the linear equation into a range of [0,1]. For squashing the predicted value from 0 to 1, we make use of the sigmoid function. It provides a solution for the classification problem that assumes that a linear combination of the observed features can be used to determine the probability of each particular outcome of the dependent variable.

E. Ensemble Method

Ensemble methods is a learning algorithm that constructs a group of classifiers and then by using the weighted vote of their predictions we classify new data points. The original ensemble method is Bayesian averaging but recent algorithms include error-correcting output coding Bagging and boosting. The various types of ensemble methods are Bootstrap AGGregating, Random Forest Models.

1. Bootstrap AGGregating

BAGGing name is given because it combines Bootstrapping and Aggregation to form one ensemble model. When a sample of data is given, many bootstrapped subsamples are taken from the sample. In each bootstrapped samples a decision tree is formed. After decision tree subsamples are formed, an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor.

2. Random Forest Models

Random Forest models will implement differentiation levels because based on different features each tree is splitted. This differentiation levels provides a greater ensemble to aggregate over, ergo producing a more accurate predictor.

IV. The Various Security Attacks Detection In Cloud By Other Authors Are Studied Below

Paper Title	Algorithm used	Security attack	Advantages	Limitations
DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments	C4.5 algorithm and decision tree	Denial of of service attack	• The article discusses about the objective of the Denial-of service attack and had proposed an DDoS model using the C.4.5 algorithm to mitigate the DDoS threat. • In this the algorithm is coupled with the signature detection techniques that generates the decision tree for detecting the DDoS attacks.	The C4.5 algorithm alone cannot detect the DDoS attack, it must be coupled with the Signature detection technique.
An Efficient Detection and Prevention of DDoS Attacks in Cloud	FireCol algorithm	Distributed denial of service	• It also discusses about the three methodologies of the Intrusion detection and demonstrates about the C.4.5 model. • Proposed about the detection and prevention of DDoS attacks in cloud environment. • The article says that internet is most popular technology and cloud computing is an internet-based computing.	In this paper existing accuracy better than proposed accuracy.
Environment			• The DDoS attacks are increasing in the cloud computing due to the essential characteristics of the cloud. It, Address the problem of DDoS attacks and present the theoretical foundation, architecture, and algorithms of FireCol. • It discusses about detection and prevention of the attack using FireCol algorithm.	

Prevent DDOS Attack in Cloud Using Machine Learning	Distributed denial of service attack	Artificial neural network	• Says about the cloud computing and its one of its security attacks that is the DDoS. • The attack is simple but it is very much powerful attack that makes resources unavailable to legitimate users. • It discusses about the prevention of DDoS attack in cloud using machine learning. • It also discusses about various machine learning algorithms. • It mainly focuses on the artificial neural network.	Here it provides methods for the detection, it mainly focuses on the detection part.
A Fuzzy Logic based Defence Mechanism against Distributed Denial of Service Attack in Cloud Computing Environment	Distributed denial of service attack	Fuzzy Logic	• The article discusses about DDoS attacks, Types of DDoS attacks, Motivation behind the DDoS attack, DDoS attack generation tools. • It says about the Defence mechanism against distributed denial of service attack in cloud computing and fuzzy based defence mechanism against distributed denial of service	In this paper it uses the trained data-based defence system but if any new type of attack happens it is difficult.
SYN Flood Attack Detection in Cloud Computing using Support Vector Machine	SYN flood attack	Support Vector Machine (SVM)	• Proposed an automated classification system for DoS attack detection in cloud computing. • They collected the network data using the Wireshark and extracted the necessary features using the Tshark. Used the Support Vector Machine for the data classification and provided with an 100% accuracy.	In this paper single attack is detected using a single algorithm, it cannot detect any other attacks.
An SVM-based framework for detecting DoS attacks in virtualized	Denial of service attack	Support vector machine and decision tree	• The article provides a svm based framework for Dos attack detection in virtualized cloud under the changing environment. • They use a filter to remove the noisy	Here they use a filter that removes the noisy data which decreases the accuracy of the detection.
Detection and Prevention of	SQL Injection attack	Fast flux monitor	• The article discusses about the SQL injection attack detection and	This paper provides detection and

SQL Injection Attack: A Survey			prevention. • Discusses about the various classical and modern SQL injection attacks. • They provide different methods and tools to detect and prevent these attacks.	prevention only for the applications.

V. Conclusions and Future Work

The Security attack detection is a very difficult problem in cloud computing. Different machine learning algorithms can be used to detect the attack but the Naïve Bayes, support vector machine (SVM), decision tree, logistic regression, and ensemble methods. Machine learning algorithms provide with the efficient output. Our future work is to enhance the system by preventing the DoS attack using the machine learning algorithm.

clouds under changing environment			data in the pre-processing step. • Their results show them that the proposed framework is better than the traditional support vector machine and the decision tree.	

References

[1] PriyankaChouhan, Rajendra Singh: Security Attacks on Cloud Computing with Possible Solution, Volume 6, Issue 1, January 2016 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering - Research Paper

[2] MarwaneZekri, Said El Kafhali, NoureddineAboutabit, and Youssef Saadi: DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments.

[3] Khalid A. Fakeeh, Ph.D.: International Journal of Applied Information Systems (IJAIS) – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 11 – No. 7, December

[4] R.T. Anitha, Dr. B. Ananthi: An Efficient Detection and Prevention of DDoS Attacks in Cloud Environment, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 7, Issue 1, January 2018, ISSN: 2278 – 1323.

[5] ZerinaMašetić, Dino Kečo, NejdetDoğru, Kemal Hajdarević: SYN Flood Attack Detection in Cloud Computing using Support Vector Machine, TEM Journal. Volume 6, Issue 4, Pages 752-759, ISSN 2217-8309, DOI: 10.18421/TEM6415, November 2017.

[6] Akshita Sharma, SarveshSingh: DDOS Attacks Detection and Prevention with Cloud Trace Back, International Journal of Innovative Computer Science & Engineering Volume 2 Issue 3; July-August-2015; Page No. 30-35.

[7] AnkuJaiswal, Chidananda Murthy P, Madhu BR: Prevent DDOS Attack in Cloud Using Machine Learning, Volume 6, Issue 6, June 2016 ISSN: 2277 128X, International Journal of Advanced Research in Computer Science and Software Engineering - Research Paper.

[8] Ayman A. A. Ali1, Prof. SaifAldeen F. Osman: Efficient DDoS Attack Detection and Prevention Framework Using Two-Level Classification in Cloud Environment, IJCSMC, Vol. 7, Issue. 8, August 2018, pg.1 – 7.

[9] N.Ch.S.N. Iyengar, Arindam Banerjee, and GopinathGanapathy, A Fuzzy Logic based Defense Mechanism against Distributed Denial of Service Attack in Cloud Computing Environment, International Journal of Communication Networks and Information Security (IJCNIS) Vol. 6, No. 3, December 2014.

[10] Adel Abusitta, Martine Bellaiche, and Michel Dagenai: An SVM-based framework for detecting DoS attacks in virtualized clouds under changing environment, Journal of Cloud Computing: Advances, Systems, and Applications Abusittaetal. Journal of Cloud Computing: Advances, Systems.

[11] Zecheng He, Tianwei Zhang, Ruby B. Lee, Machine Learning Based DDoS Attack Detection from Source Side in Cloud, 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing.

[12] Zainab S. Alwan, Manal F. Younis, Detection and Prevention of SQL Injection Attack: A Survey, International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 6, Issue. 8, August 2017.

[13] M. Alkasassbeh, G. AI-Naymat et.al," Detecting Distributed Denial of Service Attack Using Data Mining Technique," (IJACSA)International Journal of Advanced Computer Science and Applications, Vol.7, pp.436-445, 2016. Science and Information Technology, Vol 6(2), pp.1096-1099,2015.

[14] V. Hema and C. EmilinShyni," Dos Attack Detection Based on Naive Bayes Classifier," Middle –East Journal of Scientific Research 23(sensing, signal, processing, and security):398-405.

[15] Monika Malik, Dr. Yudhvir Singh, A Review: DoS and DDoS Attacks, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.6, June- 2015, pg. 260-265
[16] Rohit K Rao[1], Vasudha[2], Shraddha Bhat3, A Review on Malware Injection in Cloud Computing, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, Issue 4, April 2018.
[17] Abdul Fadil1, Imam Riadi2, Sukma Aji3*, A Novel DDoS Attack Detection Based on Gaussian Naive Bayes, Vol. 6, No. 2, June 2017, pp. 140~148, DOI: 10.11591/eei. v6i2.605.
[18] BinJia,1 XiaohongHuang,1 RujunLiu,2 andYanMa1, A DDoS Attack Detection Method Based on Hybrid Heterogeneous Multiclassifier Ensemble Learning, Hindawi Journal of Electrical and Computer Engineering Volume 2017, Article ID 4975343, 9 pages.
[19] JierenCheng ,1,2,3 ChenZhang ,1 XiangyanTang,1 VictorS. Sheng ,4 ZheDong,1 andJunqiLi1, Adaptive DDoS Attack Detection Method Based on Multiple-Kernel Learning, Hindawi Security and Communication Networks Volume 2018, Article ID 5198685, 19 pages.
[20] Wedad Alawad1, Mohamed Zohdy22, Debatosh Debnath3, A Study of DDoS Attacks Detection Using Supervised Machine Learning and a Comparative Cross-Validation, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 12, December 2017.
[21] Neha Patel and Divakar Singh, An Algorithm to Construct Decision Tree for Machine Learning based on Similarity Factor, International Journal of Computer Applications (0975 – 8887) Volume 111– No 10, February 2015.
[22] Thomas G. Dietterich, Ensemble Methods in Machine Learning, Oregon State University, Corvallis, Oregon, USA.
[23] Tzong-An Su, A Mechanism to Prevent Side-Channel Attacks in Cloud Computing Environments, Fenh Chia University Taichung, Taiwan.
[24] ShrutiPuri and ManojAgnihotri, Review: A Study on Malware Detection in Cloud Network Targeting Cloud Infrastructures, International Journal of Innovations in Engineering and Technology (IJIET).