Cloud based Security Framework for Anomaly Based Intrusion Detection using Machine Learning Techniques

Dr. Anurag Rai[1]
Director – Admin and Research,
JBIT, Dehradun, India[1]

Amit Saxena[2]
PhD Research Scholar,
UTU, Dehradun, India[2]

Dr. Manish Manoria[3]
Director, Sagar Institute of Research & Technology, Bhopal[3]

***Abstract*: Cloud is one of the most recent trends in the domain of Computing. It enables the Service Providers to have an optimal usage of their Resources in order to gain more profit out of the available resources and computing capabilities. Since its inception, it had revolutionized the concept of Service Models as it is beneficial and cost friendly for both the Customers as well as Service Providers. But as said, every scientific discovery has its own benefits as well as adverse affects; the same is applicable to Cloud also. Now, due to easier implementation and large subscribers, the traffic on Cloud systems is increasing at an alarming rate, thereby providing opportunities for Hackers and other Unauthorized Users. There are various traditional approaches for implementing Intrusion Detection Systems but in this scenario, their performance will degrade significantly due to excessive Load, high Traffic, large number of Users and Resources. In this environment, an implementation of dynamic algorithm is desirable that can handle excessive load, high resource and user count. Such dynamic algorithm can be implemented in an optimal manner using Machine Learning techniques. Intrusion Detection Systems can be classified into various types, but the most common and implementable form is Anomaly based IDS. In such systems, the behaviour of the System parameters and stake holders is being observed on continuous basis. If an entity or a stake holder is behaving differently than its observed behaviour; it indicates that something went wrong. If this modified behaviour is continuous, then it means there has been an illegal activity being performed in the system. After this, the IDS take necessary actions to handle such an alarming situation. Developing an Anomaly based Intrusion Detection System using Machine Learning technique will be a suitable solution for developing a Security Framework for Cloud environment, so that the availability, fault tolerance, scalability and reliability of the Cloud environment should remain persistent, even in case of Fault or unauthorized access.**

Keywords:- Cloud Computing, Cloud Security, IDS, Anomaly based IDS, ML for IDS

I. INTRODUCTION

Cloud Computing is one of the most recent trend and hot cake in the domain of Computing and rapidly growing Computational Model in today's world. It provides Convenient and On – Demand Network Access to a shared pool of configurable Computing Resources like Servers, Storage and Applications as a "Service" over Internet and using the underlying Network Infrastructure for fulfilling the needs of the Users. It can be considered as a hybrid of Distributed Computing, Grid Computing, Utility Computing, Network Computing and Virtualization [4].

These Services and Resources can be rapidly provisioned and released with minimal Service Provider's Interaction and in an optimal manner. There are five essential characteristics of Cloud Computing that makes it the most applicable trend in Computing in recent times. These characteristics include On Demand Self Service, Broad Network Access, Rapid Elasticity, Measured Service and Resource Pooling (Multi Tenant Model and Location Independence) [26].

The Users can access the Services and Resources offered by Cloud using Cloud Service Models or Cloud Deployment Models. Cloud Service Models are classified as "Software as a Service", "Platform as a Service" and "Infrastructure as a Service". Generally, Cloud Service Model is referred as "X as a Service", where "X" denotes a specific Service offered by the Cloud to its Users. Cloud Deployment Models are categorized as Private Cloud, Community Cloud, Public Cloud and Hybrid Cloud. Deployment Model specifies the type of Access and Reachability that Cloud Users can have in the System [26].

II. CLOUD SECURITY

Security is one of the most important aspects of Cloud Computing that needs to be taken care off at the earliest so as to achieve the optimal utilization of Resources in effective manner. The Security issues in Cloud can be identified at various levels in the architecture of Cloud like Applications, Networks, Information Storage, Virtualization, Authentication and Authorization. Solutions to these issues or threats will create a path for utilization of Cloud Computing Model in the best possible manner [13]. Customers place their vital information on Storage devices located to Cloud Data Centers and it is necessary that this information can be kept safe from unauthorized access or hacking, i. e. Confidentiality, Integrity and Availability of User Information can never be compromised [15].

Apart from these issues, there are certain Security Attacks that also require handling for effective operation of the System. An Attack is considered as an act of breaching the Security of the System through unfair or illegal or illegitimate means, thereby compromising the Computer

Security Policies, Acceptable Use Policies and Standard Security Practices. The common attacks that can be generated in Cloud Systems include Probing, Denial of Service Attack, User to Root and Remote to Local [16].

Cloud Security Controls are classified into four broad categories as Deterrent Controls, Preventive Controls, Detective Controls and Corrective Controls. The Deterrent Control Mechanisms reduces the level of Attack by generating an Alert. Preventive Control Methods strengthen the Preventive Actions against Threat or Attacks. Corrective Control reduces the severity of an Attack. Detective Controls predict and identify the type of Attack and inform Administrator to take necessary Action to handle the Attack [18].

Common solutions that can be implemented to handle Security issues include Encrypted Data Transmission inside and outside the Cloud Infrastructure, Intrusion Detection System (IDS), Firewalls and Authentication approaches. Considering the amount of Resources, Services and Users of the Cloud, IDS is the best possible solution of all for this cause [13].

III. INTRUSION DETECTION SYSTEM

Intrusion Detection System falls in the category of Detective Control Mechanisms in the Security Control of Cloud. An IDS predicts and detects the Suspicious, Malicious, Abnormal, Attacker and Intruder and informs to the Administrator, so that appropriate action can be taken to reduce the damage [18]. IDS commonly used two approaches to perform its operation, as Misuse Detection and Anomaly Detection. Misuse Detection identifies Intrusions based on Known Patterns of Malicious Activities or simply Signatures that represents a specific Threat or an Attack. On the other hand, Anomaly Detection identifies Intrusions based on deviations from Normal behavior. IDS in Cloud are classified as Host based IDS, Network based IDS and VM based IDS, specifying the location where the IDS is placed or deployed [5].

Anomaly based IDS mainly works on three main approaches, as Statistical based Technique, Knowledge based Technique and Machine Learning. It gathers the Data over a period of time regarding the operation of the System and then identifies whether there is any activity which can be considered as Malicious or Harmful using above mentioned approaches. In Statistical approach, random observations are used to represent the System behavior. It involves observing the Data currently flowing in the System and compares it against the Statistical profiles trained earlier. Knowledge based approach uses pre defined Knowledge to capture Attacks. It mainly involves the use of Expert System to perform its task. Machine Learning technique uses Models created as a base for classifying Data as Normal or Malicious / Anomalous. On initial basis, ML Models result in high overheads but as the time passes, the performance of the System improves greatly from the Data learnt during previous operations [21]

IDS can also be implemented using various techniques apart from those specified above. These techniques include Artificial Neural Networks, Fuzzy Logic, Genetic Algorithms, Association Rule, Support Vector Machines, Baye's Theorem, Naïve Baye's Classifier, Clustering approaches, Decision Trees or a Hybrid approach combining two or more techniques stated above to achieve better performance and effectiveness of the IDS [14, 16].

IV. LITERATURE SURVEY

C. Modi, D. Patel, B. Borisanya, A. Patel and M. Rajarajan in 2012 proposed a Novel Framework for Intrusion Detection in Cloud. Their approach is based on the concept of Network Intrusion Detection System (NIDS) combined with SNORT and Decision Tree. SNORT is used to detect Known Attacks and Decision Tree is used to predict that the Event is Malicious or not by observing the previously stored Network Events. SNORT is a well known Open Source Packet Sniffer used for Signature based Intrusion Detection. Their approach ensures Low FALSE Positives and High Detection Accuracy with affordable Cost. They had classified the Attacks as External and Internal. They had used NSL KDD Data Set and 10 % KDD Intrusion Detection Data Set as Training Data. Their Results shows that the Performance of the System is more than 95 % on both the Data Sets. TRUE Positive Rate is approx. 96 % and False Positive Rate of around 2 % indicating a High Detection Accuracy [4].

A. H. Bhat, S. Patra and Dr. D. Jena in 2013 proposed a Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines. They proposed an approach for Virtual Machine Monitoring based on the principle of Anomaly based Detection using Machine Learning approach. Their approach is mainly divided in 2 parts. First is deploying an Naïve Baye's Tree Algorithm for Anomaly based Intrusion Detection and Second is Anomaly Detection using a Hybrid approach for NB Tree and Random Forest Classifiers. They used KDD and NSL KDD Data Sets for Training the System. The proposed architecture comprises a Front End and a Back End. In Front End, Preprocessing and Feature Construction Components are used for both Training and Testing. In Back End, the output of Feature Construction is used for Learning and Anomaly Detection. Results show that the Performance of the System is very high with an Accuracy of approx 99 % and a False Positive Rate of approx 2 % [5].

Dr. Y. K. Sharma and D. R. Monica in 2019 proposed Deep Learning Approach for Anomaly based Intrusion Detection. In Training Phase, Genetic Algorithm is used to generate Genetic Rules and then these Genetic Rules are given as Input to Fuzzy Logic Controller. Here, probability of each Attribute is calculated which is used in later phase for Classification of Data as Normal or Attack. In Testing Phase, the Fuzzy Rules are given as Input to the Neural Network Algorithm for the Classification of Sub Attack. The System will collect the Network Traffic Data using Packet X LIB and WINCAP Driver and Neural Network

Algorithm will be applied on all instances. Transfer Function will be used to evaluate the Wright of each Node and defined Thresholds will classify the Sub Attacks. Results show that the proposed approach is having an average Detection Rate of approx 91 % and an average Error Rate of approx 7 % [6].

N. Thakkar, M. Karamta, S. Joshi and M. B. Potdar in 2019 proposed a technique for Anomaly Detection and Categorization in Cloud Environment using Deep Learning. Their approach is based on the concepts of Deep Neural Networks (DNN), Deep Belief Networks (DBN) and Restricted Boltzmann Machine (RBM). Their System comprises of various Modules for carrying out the task which include Network Traffic Data Collector Module, Data Preprocessing and Anomaly Detection Module for Attack Traffic Classification. Network Traffic Collector Module is implemented at every Network Router to capture the Incoming Traffic of Cloud Infrastructure. The Collected Data will be stored in Server and it will be passed to the next Modules. This Collected Data will be preprocessed by the Data Preprocessing Module. HADOOP and MAP – REDUCE are implemented for this operation. After Preprocessing, the Data will be forwarded to the Detection Module. Hybrid Deep Learning Method is implemented for Anomaly Detection with High Accuracy. This Module will decide whether the Data specifies an Anomaly or Normal scenario. Precision, Recall and Accuracy are parameters used for evaluating the Performance of the proposed system. Accuracy specifies Percentage of Accurate Classifications, Precision specifies Percentage of predicting Positive Anomalies over Number of Instances declared as Positive and Recall specifies Percentage of predicting Positive Anomalies over Number of Actual Positive Anomalies. Results show that the proposed System is having a Precision of approx 94 %, Recall Rate of approx 97 % and Accuracy of 96 % [8]

Z. Chiba, N. Abghour, K. Moussaid, A. E. Omri and M. Rida in 2019 proposed a New Anomaly Intrusion Detection System in Cloud Environment based on Optimized Back Propagation Neural Network using Improved Genetic Algorithm. The proposed technique detect Intruders and Suspicious Activities in and around the Cloud Computing environment by monitoring Network Traffic, while maintaining Confidentiality, Availability, Integrity and Performance of Cloud Resources and Services. The System will monitor the Traffic coming in and going out of the System. 2 Strategic Positions are specified for installation of their proposed IDS, one at Front End of Cloud and another at Back End of Cloud. IDS placed at Front End will detect Network Intrusions and Attacks coming from external Networks of Cloud and acts as a Firewall or additional Preventive Layer for Security. IDS placed at Back End will detect Intrusions occurring in Internal Network. Its main task is to mainly monitor the Virtual Machines and their behavior. Results show that the proposed approach has an Accuracy of approx 99 %, Precision of more than 99 %, False Positive Rate of less than 1 %, False Negative Rate of less than 2 %, True Positive Rate of more than 98 % and True Negative Rate above 99 % [11].

D. A. A. G. Singh, R. Priyadharshini and E. J. Leavline in 2018 proposed Cuckoo Optimization based Intrusion Detection System for Cloud Computing. They proposed the Cuckoo Optimization approach for Feature Selection. They combined Cuckoo Optimization with Naïve Baye's Classifier for selecting the significant Features from Network Data. Results show that the Accuracy of the proposed System is approx 97 % with 3 Features and average Accuracy of approx 72 % which is better if compared with traditional Naïve Baye's Classifier or other Hybrid NB Classifiers [18].

COMPARATIVE ANALYSIS

All the proposed solutions stated above used KDD and NSL KDD Data Sets for training and testing of their implemented Systems. Certain parameters are required to justify the performance of the developed Solutions. The parameters taken into consideration are True Positive, True Negative, False Positive, False Negative, Recall, Precision and Accuracy [8, 11, 19].

These parameters justify the ability and performance of the IDS. True Positive (TP) indicates the Correct Prediction of Attack, True Negative (TN) indicates the Correct Prediction of Normal Behavior, False Positive (FP) indicates the Wrong Prediction of Normal Behavior as Attack and False Negative (FN) indicates the Wrong Prediction of Attack as Normal Behavior. High TN and TP results in best possible performance by IDS, while High FN and FP reduces the performance and effectiveness of IDS. FP reduces the capability of System to detect Intrusion and FN enhances the System susceptible to Intrusion [11].

Accuracy specifies the Percentage of Accurate Classifications [8], as:

$$\textbf{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

Precision specifies the Percentage of predicting Correct Positive Outcomes over the Total Outcomes declared as Positive [8]. It will be computed, as:

$$\textbf{Precision} = TP / (TP + FP)$$

Recall specifies the Percentage of predicting Correct Positive Outcomes over the Total Actual Positive Outcomes [8], as:

$$\textbf{Recall} = TP / (TP + FN)$$

The table below gives a Comparative Analysis of the above specified approaches on the basis of the Location where the System is deployed like H – IDS, N – IDS or VM – IDS, Data Set used for Training the System like Open Source KDD or NSL – KDD and the approach used for implementing the System.

R #	IDS Type	Approach Used	Data Set
[4]	N - IDS	SNORT + Decision Tree	KDD
[5]	VM – IDS	Naïve Baye's + Random Forest	KDD
[6]	N – IDS	Hybrid (ANN + GA + Fuzzy)	KDD
[8]	N – IDS	Deep Learning (DNN + DBN + RBM)	NSL KDD
[11]	N – IDS	BPNN + GA	KDD
[18]	H – IDS	Cuckoo Optimization	Hybrid Data Set

Table 1: Comparison on the basis of type of IDS, Data Set and Approach used

The Table below compares the Performance of above specified IDS approaches on the basis of their True Positive (TP) Rate, False Positive (FP) Rate, True Negative (TN) Rate and False Negative (FN) Rate.

R #	TP Rate	FP Rate	TN Rate	FN Rate
[4]	96.25	1.91	98.08	3.74
[5]				
[6]	91.00			10.75
[8]				
[11]	98.46	0.11	99.89	1.54
[18]				

Table 2: Comparison on the basis of TP Rate, FP Rate, TN Rate and FN Rate

The Table below compares the Performance of above specified IDS approaches on the basis of Precision, Recall and Accuracy.

R #	Precision	Recall	Accuracy
[4]	99.32	96.25	96.71
[5]	99.00	99.10	99.07
[6]			
[8]	94.23	96.43	95.53
[11]	99.96		98.82
[18]			71.70

Table 3: Comparison on the basis of Precision, Recall and Accuracy

The Table below compares the various approaches specified above on the basis of certain Parameters like Technique used, Data Set, type of IDS, TP Rate, FP Rate, TN Rate, FN Rate, Precision, Recall and Accuracy.

R #	IDS Type	Approach Used	Data Set	Precision	Recall	Accuracy	TP Rate	FP Rate	TN Rate	FN Rate
[4]	N - IDS	SNORT + Decision Tree	KDD	99.32	96.25	96.71	96.25	1.91	98.08	3.74
[5]	VM - IDS	Naïve Baye's + Random Forest	KDD	99.00	99.10	99.07				
[6]	N - IDS	Hybrid (ANN + GA + Fuzzy)	KDD				91.00			10.75
[8]	N – IDS	Deep Learning (DNN + DBN + RBM)	NSL KDD	94.23	96.43	95.53				
[11]	N - IDS	BPNN + GA	KDD	99.96		98.82	98.46	0.11	99.89	1.54
[18]	H - IDS	Cuckoo Optimization	Hybrid Data Set			71.70				

Table 4: Comparison of various approaches

V. CONCLUSION

Cloud Computing is an important Model to provide Services and Resources to Users in an optimal manner. It is an On Demand and Pay per Use Model for providing Services and Resources to Users in a convenient manner. Many Users kept their Data on the Servers and Storage and it is necessary to keep that Data secure and away from unauthorized access.

Intrusion Detection System is an effective and efficient mechanism for achieving Security in Cloud environment. It ensures that the Traffic coming in and going out of the Cloud System must be observed before allowing to pass, so as to ensure that no Malicious Data penetrate the System and breach its Security Policies.

Machine Learning approaches provide an optimal way to implement Anomaly based IDS for securing the Cloud Systems. Various Machine Learning approaches are implemented in this regard like Decision Trees, ANN, Fuzzy Logic, GA, SVM and others. The benefits of ML based IDS is that after Training, it improves itself and learn new Patterns as the time goes by, thereby enhancing the effectiveness of the System. In this paper, we had compared the various Anomaly based IDS approaches and analyzed their Performance on the basis of certain Parameters. An IDS with high True Positive Rate is optimal for ensuring the Security of the Cloud System.

The Performance Analysis shows that the ML based Anomaly IDS are having an Accuracy of approx 98 %, Recall Rate of approx 97 % and Precision of approx 99 %.

We can also implement ML based Anomaly IDS using some other approaches in a Hybrid manner like Ant Colony Optimization, Swarm Intelligence, Particle Swarm Optimization, etc to enhance the Performance of the IDS.

REFERENCES

[1] P. G. Teodoro, D. J. Verdejo, G. M. Fernandez and E. Vazquez, "Anomaly based Network Intrusion Detection: Techniques, Systems and Challenges", Elsevier Journal of Computers and Security, Volume 28, Page 18 – 28, 2009

[2] A. Patel, M. Taghavi, K. Bakhtiyari and J. C. Junior, "An Intrusion Detection and Prevention System in Cloud Computing: A Systematic Review", Elsevier Journal of Network and Computer Applications, Volume 36, Page 25 – 41, 2013

[3] M. Gander, M. Felderer, B. Katt, A. Tolbaru, R. Breu and A. Moschitti, "Anomaly Detection in the Cloud: Detecting Security Incidents via Machine Learning", Springer, Page 103 – 116, 2013

[4] C. Modi, D. Patel, B. Boridanya, A. Patel and M. Rajarajan, "A Novel Framework for Intrusion Detection in Cloud", ACM, Page 67 – 74, 2012

[5] A. H. Bhat, S. Patra and Dr. D. Jena, "Machine Learning Approach for Intrusion Detection on Cloud Virtual Machines", International Journal of Application or Innovation in Engineering and Management (IJAIEM), Volume 2, Issue 6, Page 57 – 66, June 2013

[6] Dr. Y. K. Sharma and D. R. Monika, "Deep and Machine Learning approaches for Anomaly based Intrusion Detection of Imbalanced Network Traffic", IOSR Journal of Engineering (IOSR JEN), Page 63 – 67, 2019

[7] R. Jamadar, S. Ingale, A. Panhalkar, A. Kakade and M. Shinde, "Survey of Deep Learning based Intrusion Detection Systems for Cyber Security", International Journal of Research and Analytical Reviews (IJRAR), Volume 6, Issue 2, Page 257 – 261, May 2019

[8] N. Thakkar, M. Karamta, S. Joshi and M. B. Potdar, "Anomaly Detection and Categorization in Cloud Environment using Deep Learning Techniques", International Journal of Computer Sciences and Engineering (IJCSE), Volume 7, Issue 5, Page 211 – 214, May 2019

[9] V. Pandey, "A Review of Intrusion Detection Technique in Cloud Architecture", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Volume 4, Issue 1, Page 781 – 800, January 2016

[10] T. Salman, D. Bhamare, A. Erbad, R. Jain and M. Samaka, "Machine Learning for Anomaly Detection and Categorization in Multi Cloud Environments"

[11] Z. Chiba, N. Abghour, K. Moussaid, A. E. Omri and M. Rida, "New Anomaly Network Intrusion Detection Syste, in Cloud Environment based on Optimized Back Propagation Neural Network using Improved Genetic Algorithm", International Journal of Communication Networks and Information Security (IJCNIS), Volume 11, Number 1, Page 61 – 84, April 2019

[12] B. C. B. Solomon and P. J. Jayarin, "Survey on Intrusion Detection System using Machine Learning Approaches", International Journal of Engineering and Computer Science (IJECS), Volume 7, Issue 5, Page 23901 – 23907, May 2018

[13] Kiran and Dr. S. Sharma, "Enhance Data Security in Cloud Computing using Machine Learning and Hybrid Cryptography Techniques", International Journal of Advance Research in Computer Science (IJARCS), Volume 8, Number 9, Page 393 – 397, November – December 2017

[14] N. Keegan, S. Y. Ji, A. Chaudhary, C. Concolato and B. Yu, "A Survey of Cloud based Network Intrusion Detection Analysis", Springer Journal of Human Centric Computing and Information Sciences, Volume 6, Issue 19, 2016

[15] I. Avdagic and K. Hajdarevic, "Survey on Machine Learning Algorithms as Cloud Service for CIDPS", IEEE, 2017

[16] P. J. Patel, Dr. J. S. Shah and M. Patel, "Comprehensive Study on Machine Learning Techniques for IDS in Cloud Computing", International Journal on Engineering Research and Technology (IJERT), Volume 3, Issue 4, Page 827 – 830, April 2014

[17] M. Z. Abedin, K. N. Siddiquee, M. S. Bhuyan, R. Karim, M. S. Hossain and K. Andersson, "Performance Analysis of Anomaly based Network Intrusion Detection System"

[18] D. A. A. G. Singh, R. Priyadharshini and E. J. Leavline, "Cuckoo Optimization based Intrusion Detection System for Cloud Computing", International Journal of Computer Networks and Information Security (IJCNIS, MECS), Volume 11, Page 42 – 49, 2018

[19] S. Kok, A. Abdullah, N. Jhanjhi and M. Supramaniam, "A Review of Intrusion Detection System using Machine Learning Approach", International Journal of Engineering Research and Technology (IJERT, IR Publication), Volume 12, Number 1, Page 8 – 15, 2019

[20] S. M. Moorthy and M. Rajeswari, "Virtual Host based Intrusion Detection System for Cloud", International Journal of Engineering and Technology (IJET), Volume 5, Number 6, Page 5023 – 5029, December 2013 – January 2014

[21] N. M. Ibrahim and A. Zainal, "Intrusion Detection Techniques in Cloud Computing: A Review", International Journal of Computer Applications (IJCA), Volume 179, Number 12, Page 26 – 33, January 2018

[22] S. A. Repalle and V. R. Kolluru, "Intrusion Detection System using AI and Machine Learning Algorithm", International Research Journal of Engineering and Technology (IRJET), Volume 4, Issue 12, 1709 – 1715, December 2017

[23] R. Dhivya, R. Dharshana and V. Divya, "Security Attacks Detection in Cloud using Machine Learning Algorithms", International Research Journal of Engineering and Technology (IRJET), Volume 6, Issue 2, Page 223 – 230, February 2019

[24] M. Gander, B. Katt, M. Felderer, A. Tolbaru, R. Breu and A. Moschitti, "Anomaly Detection in the Cloud: Detecting Security Incidents via Machine Learning", 2012

[25] R. S. Siva Kumar, A. Wicker and M. Swann, "Practical Machine Learning for Cloud Intrusion Detection", ACM, 2017

[26] P. Mell and T. Grance, "The NIST Definition of Cloud Computing", Special Publication, National Institute of Science and Technology, US