Approach to Machine Learning for Secured Cloud Computing

Amarnath Jambhaiyanahatti Lalyanaik[1](✉), Pritam G. Shah[2], Praveen Pawaskar[3], and Vinayak B. Joshi[4]

[1] VTU, Belagavi, India
amar.rv2010@gmail.com
[2] Australian Journal of Wireless Technologies, Mobility and Security, University of Canberra, Canberra, Australia
[3] Presidency University, Itagalpur, Rajanakunte, Bangalore 64, India
praveen.pawaskar@presidencyuniversity.in
[4] S.G. Balekundri Institute of Technology, Nehru Nagar, Belagavi, India

Abstract. Machine Learning (ML) is the bigger picture of this technology driven world we live in right now. With machine learning, there are clearly a lot of benefits that we are driving in our day-to-day lives. Specifically in security there is clearly benefits of speed and accuracy that when we can bring to our applications to our infrastructure to protect the citizens that are using our applications. So, as long as the end is identified or the goal is identified and the problem is stated clearly, the machine learning can be a great way for providing the security.

In this paper, a machine learning based secure cloud computing model was proposed. This machine learning model uses the Improved Intrusion Detection and Classification (IIDC) technique. Here, the improvement is in terms of better detection and classification of malicious users compared to the complex tree based model. For this purpose, we have adopted a novel method in the machine learning process that uses the combination of past and current decisions. And it calculates a final decision by computing the majority of the all the decisions. This approach is more efficient compared to all other classic learning algorithms.

***Index Trems*:** Cloud · Cloud computing · Cloud security · Intrusion detection and machine learning

1 Introduction

Cloud computing is essentially used to deliver the various IT resources and applications as a service with the help of internet. This technology uses several individual computing nodes, storage elements and a strong network among them. It has the wide requirement ranging from an individual user to a giant company as shown in Fig. 1. The most widely used services like the email, search engines, social networking applications etc, are also hosted in the cloud. The National Institute of Standards and Technology (NIST) [1] definition of cloud computing is given as below:

D. S. Guru et al. (Eds.): ICCR 2021, CCIS 1697, pp. 79–88, 2022.
https://doi.org/10.1007/978-3-031-22405-8_6

> *"Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."*

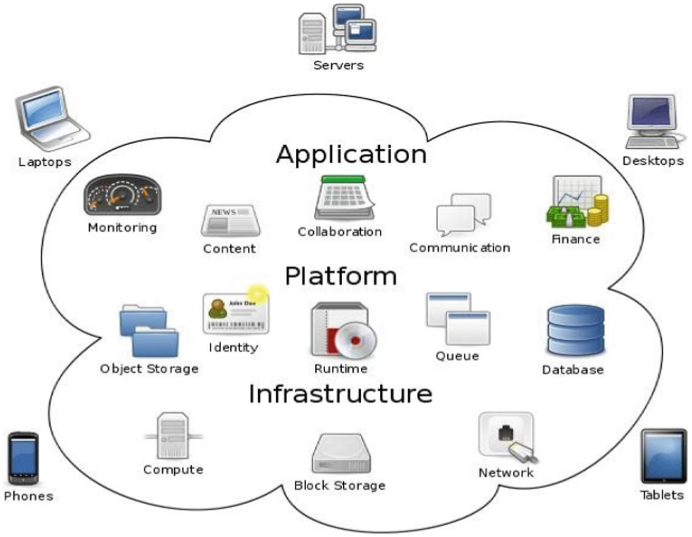

Fig. 1. Block diagram of cloud computing

Machine learning [1] for cloud security is a kind of multidisciplinary operation required to get machine learning effectively applied for cloud security. The security challenges of cloud computing environment are like detecting the anomalies and the potential security breaches. An anomaly usually logins with unknown IP address that becomes a potential security threat and some other logins which say the credentials that have been used are haven't seen before. All of these generate a bunch of independent lurks.

Machine learning is a model or we train a model with respect to particular instances so that it could predict accurate result for similar type of unknown instances. For a very longtime this had always been a challenge for various data scientist that how we could improve machine and how we could improve online security and cloud security with the help of machine learning. So they came up with two solutions number one the machine learning algorithm will classify and recognize the one which are very important and sensitive data of the user from the whole record of the user data. This machine learning algorithms scan through all the user data and classify which are sensitive user data and number two the machine learning algorithm will we would be trained to keep track of the coming threats it gives track of the user behavior so that it could predict the accurate threat that could come to end user and inform the admin at the right time before the threat could happen. So these are the two ways actually the threat coming procedure and this detection of threat procedure are actually used by previous companies like Amazon and

Microsoft. These are the companies they actually use these types of machine learning algorithm for online cloud security or internet securities. The merits are like the work will get a bit automated means a lot of manually done things will be now be automated through automation. It gives more security to the user workload as you are training a model with respect to a threat or you are training a model with respect to a user data which are sensitive or not.

2 Literature Survey

Many of the enterprises are providing the cloud based computing services and these services are increasing in day to day life. As a result potential threats are also increased in this domain. These threats are related to the security and privacy of the enterprise data. In the paper [2], the authors W. Feng, W. Yan, S. Wu and N. Liu have introduced a 2-stagemachine learning system that detects the anomalies in cloud environment. It uses the access logs of cloud based data shared services on to relationship graphs. In this approach, the machine learning methods like Odd Ball, PageRank and Local Outlier Factor to will generate the outlier indicators. Then it ensembles these indicators and introduces the wave let transform that identifies those outliers to detect the insider threat.

There exists many of the machine learning approaches to provide the cloud security. But sometimes, they may classify the nodes as the misbehaving nodes with their short-term behavioral data [3]. And they couldn't differentiate whether these misbehaving nodes are the malicious nodes or the broken nodes. This can be solved with the help of Improvised Long Short-Term Memory model. This model can learn and train the behaviour of the each user, and also it will store in the database. It can identify the misbehaving nodes as a broken node or a new user node or a compromised node. It can efficiently detect the attacks, anomaly and reduces the false alarm in the cloud networks. The performance of machine learning approaches in the prediction of cloud platforms security state will be affected by the dynamic and uncertain natures of the cloud platforms. So, by combining the internal security state and observable state of the cloud platforms, the authors Z. Li, L. Liu, Y. Zhang and B. Liu [4] have constructed the security state transition model and established a linear regression Ada Boost learning and prediction model for the observation state. He probability trend of the internal security state and its future values will be analyzed with the help of Markov model. ADOS attack detection model was implemented by Z. He, T. Zhang and R. B. Lee using machine learning approaches in the cloud environment [5]. This machine learning model makes use of statistical information from the cloud servers.

A malware detection model for cloud infrastructure was proposed by M. Abdelsalam, R. Krishnan, Y. Huang and R. Sandhu [6]. It uses the 2D and 3D convolutional neural networks and a deep learning approach. The training data is obtained from the virtual machines. 3D convolutional neural network classifiers are used to improve the accuracy. Then 2D convolutional neural network model has achieved an accuracy of 79%, whereas the 3D convolutional neural network model achieved an improved accuracy of 90%.

The machine learning based intrusion detection techniques are effectively used for cloud environment security as they are robust in learning models and due the data centric approaches. An attack feature is obtained from the network and application logs. In the paper [7], the authors N. Krishnan and A. Salim have used various machine learning models like logistic regression, belief propagation for the detection of the attack. Performance measures such as average detection time is used to evaluate the performance of the approach.

The authors M. S. Sarma, Y. Srinivas, M. Abhiram, M. S. Prasanthi and M. S. L. Ramya have proposed a novel machine learning (using KNN learning) QoS model for the effective secure wallet files classification [8]. This classifier-QoS addresses the different security issues that are related to the cloud user's community to provide the better experience for end users. Also, the authors have focussed on the barrier between the degree of trust and their implemented model.

For the effective design machine learning model, it is necessary to obtain the real-time and unbiased dataset. And these datasets are the internal and confidential matters of a particular enterprise or organization which cannot be disclosed with there searchers. As a result, the research in this problem will be limited to simulated or closed experimental environment. So, it is necessary to check the robustness of these machine learning architectures with diversified operations. So, the authors D. Bhamare, T. Salman, M. Samaka, A. Erbad and R. Jain [9], have used the UNSW dataset to train the supervised machine learning models and tested them with different dataset (ISOT). Finally, they have concluded that more research is to be carried out in this are for the validation of the machine learning model as a general solution. Machine learning based an intelligent QoS model [10] was implemented by S. Sarma, Y. Srinivas, N. Ramesh and M. Abhiram to determine the only necessary files for decryption from the entire file list in the secure wallet. This smart QoS can address the different security issues of cloud user's community.

The authors I. Aljamal, A. Tekeoğlu, K. Bekiroglu and S. Sengupta [11], have proposed a network based anomaly detection model at the cloud hypervisor level by using the hybrid algorithm. It is a combination of K- means clustering and SVM classification algorithms. It mainly improves the accuracy of detection. Distributed Denial of Service (DDoS) [12] type of attack can make the more damage to cloud environment. These attacks are a category of critical attacks that compromise the availability of the network. It uses the innocent compromised computers (called zombies) by considering their weaknesses such as known or unknown bugs and vulnerabilities for sending the bulk amount of packets to the server. Eventually, these packets will capture the huge amount of bandwidth and time. Finally, its detection model was proposed by the authors M. Zekri, S. E. Kafhali, N. Aboutabit and Y. Saadi using the machine learning techniques.

The authors A. Inani, C. Verma and S. Jain have developed an automatic data classification system for secure mobile cloud computing [13]. This system was developed using the Training dataset Filtration Key Nearest Neighbor (TsF-KNN) classification as it can classify the data depending on its confidentiality level. It has more accuracy and speed of computations than the K-NN algorithm.

The authors Z. Masetic, K. Hajdarevic and N. Dogru have proposed [14] a cloud computing threats detection and classification model based on the feasibility of machine learning models. This classification has been performed depending on the, i). Type of learning algorithm, ii). Input features iii). Cloud computing level. The obtained results can help the researchers in the selection of appropriate input features, or machine learning model, for better classification.

3 System Architecture

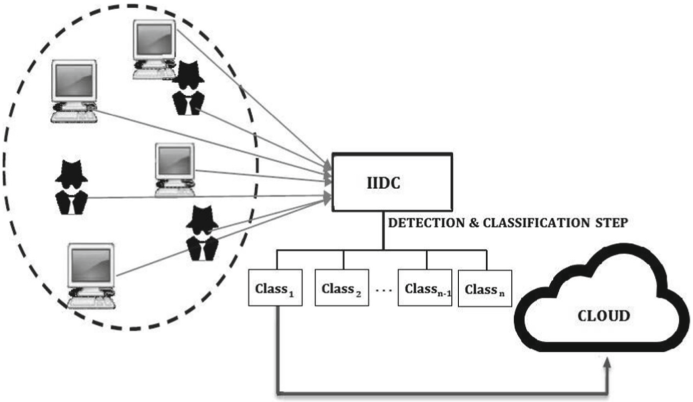

Fig. 2. Secure CC model using Machine Learning approach

The proposed block diagram of secure cloud computing (CC) model based on machine learning is shown in Fig. 2. The heart of the proposed model is the Improved Intrusion Detection & Classification (IIDC) module. The main advantage of this module is that; it uses the machine learning framework for the efficient detection and classification of intrusions in the cloud network, thus providing the high level of security for the cloud computing environment. We have considered *N* number of nodes in the above network. And each node represented with a feature vector that represents the characteristics of transmitted data packets (such as IP destination, transmission protocol). Then each output class w.r.to its feature set is represented with. The training dataset *TD* is modeled as given in the Eq. (1):

$$\begin{gathered} \forall i \in \{1, N\} \\ TD = \{\} \end{gathered} \tag{1}$$

3.1 Improved Intrusion Detection & Classification (IIDC)

The Fig. 3 shows the IIDC machine learning framework that consists of following important phases.

a. Learning phase
b. Decisions storing phase
c. Combined decision phase

3.1.1 Learning Phase

It is the combined stage of the machine training and model creation. It generates the modeling function $f(x)$ of input features (x) for each node to an output class y with the help of dataset provided to it.

$$y^i = f(x^i). \tag{2}$$

3.1.2 Decisions Storing Phase

For each received packet at time t, the developed model of the learning phase outputs a new decision classification and stored in the database. For each node i, the history of decisions are stored as follows:

$$\mathrm{H(d)} = \mathrm{i_1t_1} + \mathrm{i_2t_2} + \mathrm{i_3t_3} + \ldots + \mathrm{i_nt_n} \tag{3}$$

3.1.3 Compute Majority Decisions Phase

The procedure for computing the majority of decisions is as follows:

This module determines the node using the set of features. Then it accesses the decision history database to extract the vector of that particular node i. Now, it combines the D^{i} with its current decision. Finally, it computes the majority of decisions for the selection of best decision.

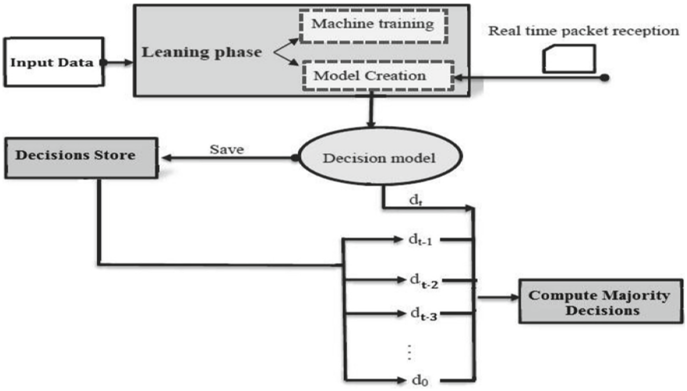

Fig. 3. Machine learning framework

4 Results

Table 1 shows the UNSW dataset that consists of normal traffic data packets as well as 9 types of attacks. We can use it as the anomaly detection models. It consists of training

(1,75,341 sets) and testing datasets (82,332 sets) used for model creation and testing purposes respectively.

Complex tree [15] from decision tree is used with 50 as the total number of nodes which are distributed proportionately with number of classes.

Accuracy Definition:

> *"The accuracy is the ratio of correct predictions over the total number of the packets in the testing set"*

Where TP → True Positive and
TN → True Negative

Table 1. Classes notation

Number	Class
1	Normal
2	Analysis
3	Backdoor
4	DoS
5	Exploits
6	Fuzzers
7	Generic
8	Reconnaissance
9	Sellcode
10	Worms

According to the obtained simulation results, the accuracy of IIDC increases with respect to time where as the accuracy of complex tree is fixed to 69% as shown in Fig. 4. It tells that the IIDC is sensitive to the time. Here, the past performance of nodes is considered in IIDC for the classification to increase the accuracy. Also, the accuracy is increased 24% more than the complex tree. Hence, the IIDC detects better the traffic anomaly than complex tree.

The detection performance of IIDC at different time indexes is shown in Fig. 5. The IIDC and the complex tree models have same performance at time index equals 1. An important note is the normal node average decision is equal to 0 and the average for malicious node is 1. In the figure; the complex tree couldn't classify the packets of the classes (2, 3, 4, and 10) because they do not have sufficient number of dataset packets for self training which has an impact on the detection performance of IIDC. Also, the difference between complex tree and IIDC performances was increasing with time.

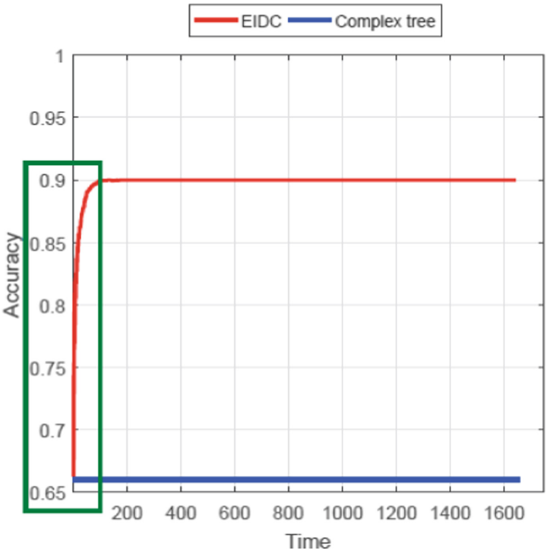

Fig. 4. Comparison of accuracies of the IIDC with complex tree w.r.t to time

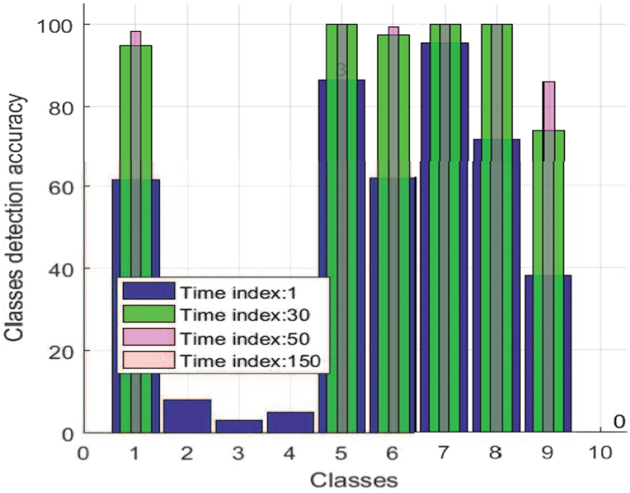

Fig. 5. Detection performances of the IIDC per class w.r.t. to time.

5 Conclusion

Hence, a machine learning based secure cloud computing model was proposed and implemented. This machine learning model provides the improved intrusion detection and classification (IIDC) module. Here, the improvement is in terms of better detection and classification of malicious users compared to complex tree based model. For this purpose, we have adopted a novel method in the machine learning process that uses the combination of past and current decisions. And it calculates a final decision by computing the majority of the all the decisions. This approach is more efficient in terms of attack detection compared to all other classic learning algorithms. Also, it has increased the classification accuracy from 66% to 90%.

References

1. Wani, A.R., Rana, Q.P., Saxena, U., Pandey, N.: Analysis and detection of DDoS attacks on cloud computing environment using machine learning techniques. In: 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, pp. 0870–0875 (2019)
2. Feng, W., Yan, W., Wu, S., Liu, N.: Wavelet transform and unsupervised machine learning to detect insider threat on cloud file-sharing. In: 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Beijing, pp. 0155–0157 (2017). https://doi.org/10.1109/ISI.2017.8004896
3. Nathezhtha, T., Yaidehi, V.: Cloud insider attack detection using machine learning. In: 2018 International Conference on Recent Trends in Advance Computing (ICRTAC), Chennai, India, pp. 060–065 (2018). https://doi.org/10.1109/ICRTAC.2018.8679338
4. Li, Z., Liu, L., Zhang, Y., Liu, B.: Learning and predicting method of security state of cloud platform based on improved hidden Markov model. In: 2018 3rd International Conference on Smart City and Systems Engineering (ICSCSE), Xiamen, China, pp. 0600–0605 (2018)
5. He, Z., Zhang, T., Lee, R.B.: Machine learning based DDoS attack detection from source side in cloud. In: 2017 IEEE 4th International Conference on Cyber Security and Cloud Computing (CS Cloud), NewYork, NY, pp. 0114–0120 (2017)
6. Abdelsalam, M., Krishnan, R., Huang, Y., Sandhu, R.: Malware detection in cloud infrastructures using convolutional neural networks. In: 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA, pp. 0162–0169 (2018)
7. Balakrishna, K., Rao, M.: Tomato plant leaves disease classification using KNN and PNN. Int. J. Comput. Vision Image Process. **9**(1), 51–63 (2019). https://doi.org/10.4018/IJCVIP.2019010104
8. Sarma, M.S., Srinivas, Y., Abhiram, M., Prasanthi, M.S., Ramya, M.S.L.: KNN file classification for securing cloud infrastructure. In: 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, pp. 05–09 (2017)
9. Bhamare, D., Salman, T., Samaka, M., Erbad, A., Jain, R.: Feasibility of supervised machine learning for cloud security. In: 2016 International Conference on Information Science and Security (ICISS), Pattaya, pp. 01–05 (2016)
10. Sarma, M.S., Srinivas, Y., Ramesh, N., Abhiram, M.: Improving the performance of secure cloud infrastructure with machine learning techniques. In: 2016 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, pp. 078–083 (2016)
11. Aljamal, I., Tekeoğlu, A., Bekiroglu, K., Sengupta, S.: Hybrid intrusion detection system using machine learning techniques in cloud computing environments. In: 2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA), Honolulu, HI, USA, pp. 084–089 (2019)
12. Zekri, M., Kafhali, S.E., Aboutabit, N., Saadi, Y.: DDoS attack detection using machine learning techniques in cloud computing environments. In: 2017 3rd International Conference of Cloud Computing Technologies and Applications (Cloud Tech), Rabat, pp. 01–07 (2017)

13. Inani, A., Verma, C., Jain, S.: A machine learning algorithm TsF K - NN basedon automated data classification for securing mobile cloud computing model. In: 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, pp. 09–013 (2019)
14. Masetic, Z., Hajdarevic, K., Dogru, N.: Cloud computing threats classification model based on the detection feasibility of machine learning algorithms. In: 2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, pp. 01314–01318 (2017)
15. Chkirbene, Z., Erbad, A., Hamila, R.: A combined decision for secure cloud computing based on machine learning and past information. In: 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marakesh, Morocco, pp. 01–06 (2019)