A SURVEY OF CLOUD NETWORK ANALYTICS

Shahida K K, NamyaMusthafa
Masters in Computer Science & Engineering
RCET, Akkikavu
shahidajubil@gmail.com,
namyamusthafa@gmail.com

Nowshad M U
Assistant professor
RCET, Akkikavu
nowshad@royalcet.ac.in

***Abstract*—Today, Cloud Computing is the well-liked choice of every organization since it provides flexible and pay-per-use based services to its users. However, the security and privacy could also be a serious hurdle in its success thanks to its open and distributed architecture that is vulnerable to intruders. This paper addresses the safety risks and challenges and analyzes the available measures to detect the anomalies in cloudenvironment.The existing IDS deployed in traditional internet or intranet environments lack the features of scalability and autonomic self-adaptation . Moreover they're not deterministic which make them unsuitable for cloud based environments. This paper analyses various techniques utilized in anomaly detection and results in machine learning is the better method which can be used for cloud network analysis.**

***Keywords*—Cloud security, ML, Anomaly detection, IDS**

I. INTRODUCTION

Cloud computing could even be a technology, where we'll store and retrieve of our data at anytime and anywhere with the utilization of internet. Now during this contemporary technology many companies and organizations are using cloud computing to store their valuable and important data. But cloud computing creates numerous security issues. the foremost problem is that user has most sensitive data. Here user must remember that each one data given to cloud provider leave the own control. When deploying this data into clouds, the cloud provider has more control on data. If an attacker has access to the cloud storage component is able to alter data in the storage. This could be done once, multiple times, or continuously. An attacker that also has access to the processing logic of the cloud also can modify the functions and their input and output data[1].

Different types of attack, as well as various attack goals, make it difficult for security administrators of cloud providers to implement optimal security solutions needed for their clients [2]. This is often because different attacks are related to different threats within the cloud computing systems, where the importance of threats varies supported the safety requirements of various clients using the cloud services.

There is various kinds of anomalies in the cloud environment; Point Anomalies, Contextual Anomalies and Collective Anomalies. Point anomalies are when a private data instance deviates from its normal activity or form it's said to be anomalous, because other data are normal. This shows that anomalous activities lie outside the boundaries of the normal region. one instance of data is anomalous if it's too away from the rest . Business use case: detecting mastercard fraud supported "amount spent." Contextual anomalies: The abnormality is context specific. this sort of anomaly is common in time-series data. The contextual anomalies occur when the occurrence of knowledge is or shows traces of anomalous character during a particular or precise context, which is that the unwanted behavior of activities that surrounds a private data instance. Only during this case it's contextual anomaly (also mentioned as conditional anomaly). Collective anomaly is when related data instances collected acts as anomalous or show unwanted activities associated with the whole data set. collective anomaly requires a relationship between data instances; spatial, sequential and graph data to cause a collective anomaly. But their occurrences as a whole or collection are often anomalous. If some linked objects are often observed against other objects as anomaly. Individual object can't be anomalous during this case, only collection of objects.

II. ANOMALY DETECTION METHODS

In the cloud networks, there are different techniques or methods that are utilized within the detection of anomalous activities; these include Threshold detection, statistical analysis, Rule-based measures, neural networks, genetic algorithms, data processing and machine learning.

A. Statistical Anomaly Detection Systems

This method of anomaly detection in cloud base network detects anomaly by observing computations within the network and creates a profile which keeps or stores the generated value in resenting their behavior. In identification of anomaly using this system , there are two profiles created; the primary one stores the traditional or anomaly rules or signatures while the other updates at regular intervals. During the update anomaly scores are calculated. If the edge value is less than the present anomaly profile generated, then it's known to be anomalous and detected. There's high probability of occurrence of normal data instances in dense regions of the model, while irregularities is seen in the low possibility regions.

Jordan Hochenbaum propose novel statistical learning-based techniques to detect anomalies in the cloud [3]. These techniques are often wont to automatically detect anomalies in statistic data of both application metrics like Tweets Per Sec (TPS) and system metrics like CPU utilization etc. Specifically, we propose the following: Seasonal ESD (S-ESD): This technique employs time series decomposition to determine the seasonal component of a given time series. S-ESD then applies ESD on the resulting time series to detect the anomalies. Seasonal Hybrid ESD (S-H-ESD): In the case of some time series (obtained from production) we encountered a relatively high percentage of anomalies. To address such cases, including the very fact that mean and standard deviation (used by ESD) are sensitive to an outsized number anomalies [4], we extended S-ESD to make use of the robust statistics median and median absolute deviation (MAD) to detect anomalies. Computationally, S-H-ESD is more expensive than SESD but is more robust to a higher percentage of anomalies. The efficacy of both S-ESD and S-H-ESD was evaluated using core metrics such as Tweets Per Sec (TPS), system metrics such as CPU and heap usage and application metrics. The evaluation was carried out from three different perspectives, viz., capacity engineering (CapEng), user behavior (UB), and supervised learning (Inj). Overall, S-H-ESD outperformed S-ESD, with F-Measure increasing by 17.5%, 29.5% and 0.62% for CapEng, UB, and Inj respectively. In light of the fact that S-H-ESD more computationally expensive than S-ESD (recall that the former requires sorting of the data), it is recommended to use S-ESD in cases where the time series into account is large but with a comparatively low anomaly count.

Now, statistical methods aren't included, because their performance heavily relies on certain data distribution assumptions. Skilled attackers can train a statistical anomaly detection to easily accept abnormal behavior as normal. All behaviors can't be modelled using statistical methods. It also can be difficult to work out thresholds that balance the likelihood of false positives and false negatives.

B. Data Mining Based Anomaly Detection Systems

The analyzing or extracting knowledge of huge data set to fine patterns that are useful to the info owner is understood as data processing . This technique uses the classification, clustering and association rule mining methods within the detection of anomalies in cloud environment. An analyst mechanism is within the data processing technique that detects anomaly by differentiating between normal and abnormal activities within the cloud. This is accomplished by stating or delineating some boundaries for valid and normal activities within the cloud network. There's also another level of focus during this technique for anomaly detection. Data mining techniques are more flexible and simply to deploy at any point.

Data mining techniques for intrusion detection proves to be better choice for both misuse detection and anomaly detection. Data mining techniques are helpful in detecting totally new attack and derivative of documented attacks. [5] reviews various cloud intrusion detection systems that uses data processing techniques for attack detection.. Shikha Agrawal, etal(2015) have done a detailed survey on various fraud detection techniques that has been carried out using data mining techniques. They have defined Clustering based Anomaly Detection techniques, Classification based anomaly detection and Hybrid approaches.

The downside during this is often that if the clients aren't informed of the knowledge that's been collected and used for mining, there's a violation of their privacy and it's illegal. There are kinds of issues available in processing detection in cloud based networks which are the priority replacement of preserving privacy and setting the wrong parameters of these privacy settings while using different rules and strategy to strengthen cloud network security.

C. Adaptive Anomaly Detection Systems

In cloud networks, possible failures or anomaly which are detected by cloud operators are detected by the Adaptive Anomaly Detection Systems using the performance data of the cloud service. The AAD detection systems utilize or maximize the log of the detected failure records that are sent in by the cloud operators to identify new kinds of failures subsequently. The AAD detection algorithm changes its behaviour by repeatedly learning from the new certified results or detection from the cloud provider soon be prepared for future detections. The failure detector in achieves 67.8% sensitivity within the experiments.

AAD detects possible failures supported the cloud performance data which are verified by the cloud operators. they're confirmed as either true failure with either failure types or normal states. Meanwhile, it draws on the observed but undetected failure records reported by the cloud operators are considered to identify new kinds of failures. Statistical methods' performance heavily relies on certain data distribution assumptions so as that they are not included. Support vector-based approaches, like support vector machines (SVM) and 1- class SVM, work well. We explore support vectors to elucidate cloud performance data, detect anomalies, and adapt the failure detector when verified detection results are available. The drawback of this method is that the accuracy of anomaly detection, there's misdetections in conjunction with proactive and reactive failure.

D. Machine Learning Based Anomaly Detection Systems

The ability for programs or software to enhance performance of their task over time by learning is a crucial technique within the detection of anomaly. Verified values or normal behavior of data are stored, when anomaly occurs or is being detected the machine learns its behavior, stores the new sequence or rule [6]. This technique creates a system which will improve on performance of the program by leaning from the previous results. The interesting part during this technique is that upon improving of performance from previous results, new information are extracted and if it requires a change within the strategy of execution to enhance performance it's done on the idea of the new information from the previous results. There are different categories of machine leaning based anomaly detection such as; Bayesian Network, Genetic Algorithm, Neural Network etc.

III. MACHINE LEARNING TECHNIQUES

With the aim of simulating the operation of human brain, neural networks adopted within the field of anomaly detection. Neural network detection approach has been employed to create prediction model, to spot intrusive behavior of traffic patterns, etc. Basically it operates in two steps. First, a neural network is trained on the traditional training data to find out the traditional class or classes. Second, each test instance is provided as an input to the neural network to see whether it's normal or anomalous. Neural Networks has the potential to enhance on data that's not complete to make a possible to detect and understand patterns that aren't visible. The Neural network does not only detect previous attacks but also unseen behaviour or patterns. Genetic Algorithms employs the evolutionary algorithm techniques like mutation, selection etc. their different process is predicated on collected rules from the knowledge on the network analysis administered by the IDS.

Problem to classify more than two classes for network anomaly detection system using machine learning techniques is discussed in [7].A model of an Online Average One Dependence Estimator (AODE) algorithm for multi-classification of UNSW-NB15 dataset built to overcome the issues in a network system. It focuses on online network anomaly-based where the machine learning approach is employed to learn the classes of UNSW-NB15 dataset either it is a normal data or various anomalous data. Online learning enable the classifier model to continuously update the features where each data instance arrived included during training phase [8]. Online AODE algorithm for multi-classification the percentage of accuracy is more than 90% except for Exploits type which is equal to 89.81%.

To detect and stop attacks, a characterization of the matter by defining different scenarios depending if we've valid and/or attack data available for training. Propose two solutions: first one is multi-class approach for the scenario when valid and attack data is available; and second solution is a one-class solution when only valid data is at hand [9]. The machine learning techniques can improve the detection capabilities of MODSECURITY in terms of the reduction of false positives and the increment of true positives. Also provided a characterization of the problem by identifying different scenarios depending on the availability of training data. The scenarios vary from the rare, but best case, where have a dataset with real application traffic to more practical scenarios where have only valid requests to an application that could be collected, for instance, during the functional testing phase. Although a low rate of false positives was obtained, the performance on attack detection decreased.

Naive bayes Rule is that the basis for several machine-learning and data processing methods. This algorithm is employed to make models with predictive capabilities [10]. It provides new ways of exploring and understanding the data. Generally, Naive bayes classifier technique is employed when the info is high and when the attributes are independent of every other. Naive bayes classifier algorithm is employed to model normal and suspicious network activity.

Several industries in many different domains are looking at deep learning as a way to take advantage of the insights in their data, to enhance their competitiveness, to open up novel business possibilities, or to resolve problem thought to be impossible to tackle. The large scale of the systems where deep learning is applied and therefore the need of preserving the privacy of the used data have imposed a shift from the normal centralized deployment to a more collaborative one [11].

Security and privacy are starting to be considered extremely important as machine learning, especially its deep formulation, is starting to be applied to industrial and critical context. The literature on this topic is starting to be characterized by several proposed solutions; however, it is equally important to have a wise use of the energy since many of these DL solutions are run on resource constrained devices. Cryptographic primitives are typically used for this scope, but are costly.

IV. CONCLUSION

Anomaly detection in could networks may be a wide area of research, and it holds an honest number of developments and proposing of detection systems. Anomalous activities occur always in cloud based or non-cloud based networks. With the different types of methods or techniques in anomaly detection in cloud based network, detection of unwanted behaviour can be traced, detected, stopped. These techniques have their limitations that create a niche between their performance metrics. In this paper, discussed the importance of anomaly detection system in cloud environment, its types, methods, and the limitations that each method is faced with such as, false alarm being created; detection accuracy is hinged on the idea of previous collected information on anomalous behavior; longer is required within the identification of attacks etc.

REFERENCES

[1] J-M Bohli, N. Gruschka, M. Jensen, “Security and Privacy-Enhancing Multicloud Architectures”, IEEE Transactions on Dependable and Secure Computing, April 2014.

[2] K. Dahbur, B. Mohammad, and A. B. Tarakji, “A Survey of Risks, Threats and Vulnerabilities in Cloud Computing,” in Proc. of the International Conference on Intelligent Semantic Web-Services and Applications (ISWSA 2011), 2011.

[3] JordanHochenbaum Owen S. VallisArunKejariwal , “Automatic Anomaly Detection in the Cloud Via Statistical Learning”, april 2017.

[4] Christophe Leys, Christophe Ley, Olivier Klein, Philippe Bernard, and Laurent Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 2013.

[5]Pinki Sharma&JyotsnaSengupta, “Intrusion Detection Using Data Mining in CloudComputing Environment”, international Journal of Distributed and Cloud Computing, December 2018

[6]Arif Sari, “A Review of Anomaly Detection Systems in Cloud Networks and Survey of Cloud Security Measures in Cloud Storage Applications”, Journal of Information Security, Vol.06, April 2015.

[7]MukrimahNawir&Amiza Amir, “multi-classification of unsw-nb15 dataset for network anomaly detection system”, Journal of theoretical and applied information technology, august 2018.

[8] F. Gumus, C. O. Sakar, Z. Erdem, and O.Kursun, “Online Naive Bayes classification for network intrusion detection”, ASONAM2014 - Proc. 2014 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal ,2014.

[9] Gustavo Betarte&Rodrigo Mart´ınez, “Web Application Attacks Detection Using Machine Learning Techniques”, 17th IEEE International Conference on Machine Learning and Applications, 2018.

[10] Shubhangi S. Gujar& B. M. Patil, “Intrusion Detection Using Naïve Bayes for Real Time Data”, International Journal of Advances in Engineering & Technology, May 2014.

[11] Christian Esposito&Shadi A. Aljawarneh, “Securing Collaborative Deep Learning in IndustrialApplications within Adversarial Scenarios”, Ieee Transactions on Industrial Informatics, 2018.

[12] Tara Salman, Raj Jain, Mohammed Samaka, , Machine Learning for AnomalyDetection and Categorization in Multi-cloud Environments IEEE 4th International Conference on Cyber Security and Cloud Computing, Volume: 25 , Issue: 3 , March 2016.

[13] Armstrong Nhlabatsi, Jin B. Hong, Dong Seong Kim, Rachael Fernandez, AlaaHussein, NooraFetais, and Khaled M. Khan, Threat-specic Security Risk Evaluation in the Cloud, IEEE transactions on cloud computing, October 2018.

[14] Farhan Bashir Shaikh, SajjadHaider, Security Threats in Cloud Computing, 6th International conference on Internet Transaction and Secure transactions, Abu Dhabi, United Arab Emirates, 14 December, 2011.

[15] Anna L. Buczak, ErhanGuven ”A survey of data mining and machine learning methods for cyber security intrusion detection”, IEEE communications surveys and tutorials, 2016.

[16] Matthias Gander, Basel Katt, Michael Felderer. ”Anomaly detection in the cloud: detecting security incidents via machine learning”, Conference paper in communications in computer and information science January 2013.

[17] Mustafa Efendioglu, AlperSen, YavuzKoroglu, ”Bug Prediction of SystemCModels using Machine Learning”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2018.