ANOMALY BASED INTRUSION DETECTION SYSTEM IN CLOUD COMPUTING

[1]Prajakta Hande , [2]Shrest Garian , [3]Satyam Tiwari , [4]Babita Bhagat
UG Student[1,2,3], Assistant Professor[4]
Department of Computer Engineering,
Pillai HOC College of Engineering and Technology, Rasayani,Raigad,India

Abstract : *Nowadays in the cloud networks immense amounts of data are being worked upon and stored and transmitted from one database to another. The data that is transmitted from one database to another is bound to be exposed to attacks. Although multiple methods or softwares are available to secure data, alternatives exist. Various hybrid approaches have been made in order to detect known and unknown intrusion attacks more precisely. The Anomaly Detection System is one of the available Intrusion Detection techniques. It's a sector in the cloud environment that is been developed for the detection of unusual activities in the cloud networks. Although, there are multiple Intrusion Detection techniques available in the cloud environment, this paper focuses on developing an IDS technique in cloud networks through different classification and conducts relational study on the log files of different databases, to bring into spotlight their strength and weakness in terms of security. It changes the execution measurements into a low-rank framework and after that ascertain the orthogonal separation utilizing the Robust PCA calculation. The proposed display refreshes itself recursively learning and changing the new edge esteem so as to limit remaking mistakes.*

IndexTerms **- Cloud Computing, Infrastructure as a Service, Platform as a Service, Software as a Service, Robust Principal Component Analysis, Multivariate Adaptive Statistical Filtering ,Entropy-based Anomaly Testing , Automatic Anomaly Detection ,Singular Value Decomposition, etc.**

I. INTRODUCTION

Cloud computing isn't giving any affirmation on information security however it is a prerequisite in the IT world. The cloud computing innovation permits the customers for substantially more dependable and proficient figuring by brought together capacity, memory, preparing and data transfer capacity. This permits the cloud employments adaptability in getting to the cloud information over the cloud network. The stream of system that is been made by cloud computing framework demonstrates client's conduct in beneficial operation or utilization. The cloud computing condition has across quantities of security challenges. The greater part of them has been tackled up to a degree, other security angles jump up and it's key to know before associations switch completely. Anomaly detection is a step by step process of locating the patterns in a dataset whose behavior is unusual on expected. These unexpected behaviors are also termed as outliers. The anomalies cannot always be classified as an attack but it can be a abnormal behavior which is previously unknown. It is not necessarily harmful. The anomaly detection provides very note worthy and critical information in various fields, for example Credit card thefts or identity thefts. When data has to be checked in order to predict known or unknown data mining techniques that are being used. These include clumping, classification and machine based learning techniques. Multiple approaches are also being created in order to attain higher level of accuracy on finding abnormalities. In this approach the authors try to combine existing data mining algorithms to derive better results. Thus detecting the unusual or unexpected behavior or anomalies will result in study and categorization of it into particular type of intrusions. In cloud networks, traffic or transferred data comes from multiple domains. There's a rapid change that can be seen in the cloud systems due to the patterns or behavior of clients/tenants using the cloud infrastructure and the state of the unprotected services. In cloud environment, numerous challenges of detecting anomaly such as misconfiguration or high ratio of legitimate traffic in the network. The importance of the anomaly detection in cloud networks is the unacceptable activities in data that brings the importance of reason for such anomaly in the information. Generally, the mostly used or commercial ready-made systems for detecting anomalies are based on signatures or rules. We propose an automated anomaly detection technique in a distributed environment to detect anomalies in the cloud infrastructure. An adaptive algorithm is introduced which uses reconstruction errors to determine the sample size and update the threshold value. The effectiveness of our methodology is evaluated using cluster traces and Anomaly datasets.

II. PROBLEM SATATEMENT

It requires some system automation to screen such a huge framework. This thus expects checking to pick up experiences into the operation of equipment, frameworks, and applications running in the cloud. Checking of the framework is the key component for following framework behavior and distinguishing anomalous behavior. Anomalous behavior is recognized as irregular behavior because of execution issues, disappointments, and setup issues. The variation from the norm can cause startling behavior and result in wastefulness and downtime of the datacenter. Current datacenters comprise of thousands of virtual machines which require dynamic

asset planning to work proficiently and cost-adequately. These server farms need to meet the shifting interest for different assets, e.g. CPU and memory; the scheduler must apportion or re-distribute the assets progressively. This requires observing of asset use in request to identify irregular behavior. It is basic to watch the server measurements (e.g., latency, CPU, memory), spoke to time arrangement, for any unordinary conduct. Early identification of these anomalous time arrangement is basic for taking pre-emptive activity to secure clients and give a superior client encounter.

III. EXISTING SYSTEM

Existing technique uses the classification, clustering and association rule mining methods in the detection of anomalies in cloud environment. An analyst mechanism is in the data mining approach that detects unusual patterns by differentiating between normal and abnormal activities within the cloud. This is accomplished by stating or tracing some boundaries for valid and normal activities in the cloud network. There's also an additional level of focus in this technique for anomaly detection. Customers or users of a cloud service have to pay for the data mining tool that's being used. Data mining is used by Cloud Service Provider (CSP) to provide service for their users or clients using their cloud service. The downside in this is that if the clients are not informed of the information that's been collected and used for mining, there's a encroachment of their privacy and it's illegal. There are varieties of issues available in data mining detection in cloud based networks which are the priority replacement of preserving privacy and setting the wrong parameters of these confidentiality settings while using different rules and strategy to heighten cloud network security.

IV. PROPOSED SYSTEM

We will implement a versatile anomaly detection system which researches principal components of execution measurements. It changes the execution measurements into a low rank lattice and after that compute the orthogonal separation utilizing the Robust PCA algorithm. The proposed demonstrate refreshes itself recursively learning and altering the new edge esteem keeping in mind the end goal to limit recreation blunders. We utilize a self versatile based anomaly detection technique to distinguish unusual practices. Our technique investigates the log documents and predicts asset use to make marked anomalies. Recreation blunders help to alter the limit esteem which distinguishes the abnormalities proficiently. Our strategy comprises of 5 stages: (1) pre-processing , (2) metric collection , (3) feature extraction , (4) prediction and (5) anomaly detection. We present a productive and conveyed anomaly detection algorithm that experiences gentle presumptions on uncorrupted focuses, which recuperates the ideal low-dimensional subspace and recognizes the defiled focuses. Our method includes network deterioration utilizing SVD (Singular Value Decomposition). The PCA is changed into a low-rank estimate to information network utilizing the lower dimensional approximating subspace utilizing SVD. The anomalous data are typically with a higher extent and fluctuation in the projection plane.

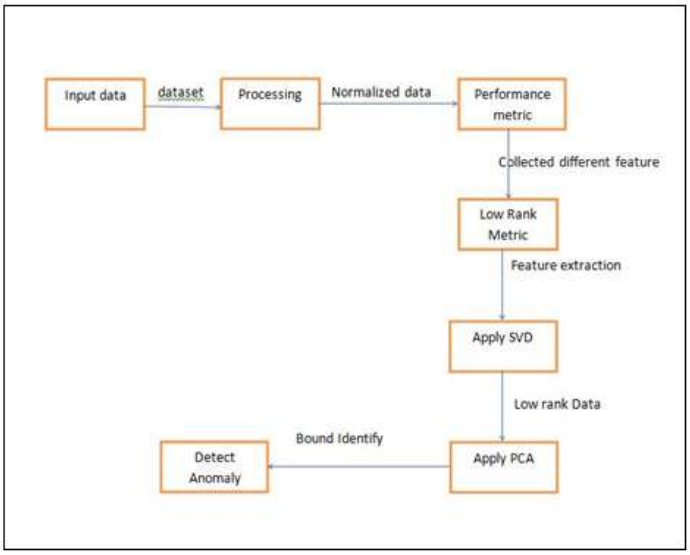

Fig -4.1: Proposed System Architecture

V. SYSTEM MODULES

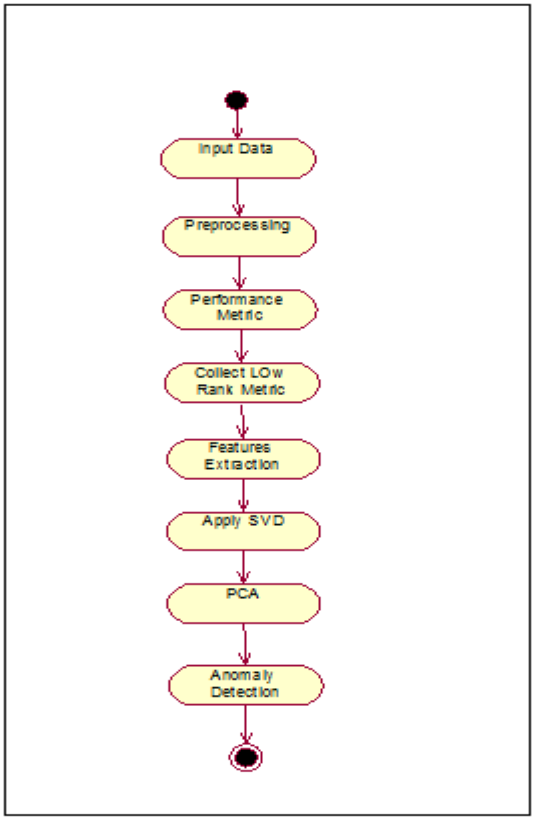

Fig -5.1: Work Flow of System

1] Preprocessing

The Input file is pre-processed and transformed into a form that can be readable through our model. The data from each node is preprocessed and transformed into a form that can be easily accessible by the model. Each sample value is transformed to a normalized form by dividing the sample value by the mean of all the samples. Once normalization is completed, each of the normalized sample values is binned with each other.

2] Metric Collection

There are different metrics available in the data, and it is hard to identify the right metrics so, the metric selection is a necessity for data analysis. We collect a uniform set of metrics from the nodes and concatenate them into one matrix, it is necessary to select an optimal subset of metrics. Metric selection is also known as dimensional reduction. The data presented in a low dimensional subspace is easier to separate into different classes. To collect different features we aggregate the metrics in per second intervals for given time frames.

3] Feature Extraction

In order to detect anomalous behavior in resource utilization for large-scale industrial distributed systems such as cloud computing and data warehouses, we need to predict the consumption. PCA is the most widely used analytical technique for data analysis and magnitude reduction today. However, it is breakable with respect to the depraved input data matrix which often threatens its validity. PCA uses the SVD (Singular Value Decomposition) to find low-rank representations of the data. The robust version of PCA (RPCA) recognizes a low-rank representation and a set of exceptions by regularly calculating the SVD and applying thresholds to the singular values and errors for each step of calculation.

4] Anomaly Detection

The Robust PCA technique method compares observed CPU utilization with a lower bound and higher bound threshold. When the utilization goes above the higher bound threshold, an abnormal behavior is detected, and this method will raise an alarm as an abnormal activity. To obtain the near optimum threshold value, we load the complete historical data of the CPU utilization to our model. From the historical trace, we calculate the lower and upper threshold bound representing abnormal behavior outside the acceptable range. The lower bound is 0 as there are 0 processes running. Other than the highest values, the corresponding 1% boundary values are chosen as the upper bound thresholds. It also shows the result when applying our model to the CPU usage for anomaly detection. The models ability to predict anomaly precisely is evaluated by four metrics: precision, recall, false positive rate and F-measure.

VI. CONCLUSION

Large-scale and complex cloud computing systems are susceptible to failures, which can significantly affect the cloud dependability and performance. We collected all the performance traces and normalized it. Robust PCA is used to transform original data into low-rank representation using recursive SVD and soft threshold. A self-adaptive threshold technique is used. The threshold is updated during the learning phase. RPCA uses an efficient approach to decompose into low-rank representation.

REFERENCES

[1] H. S. Pannu, J. G. Liu, Q. Guan, and S. Fu, "AFD: Adaptive Failure Detection System for Cloud Computing Infrastructures", In: Proceedings of 31st IEEE International Performance Computing and Communications Conference (IPCCC), (2012), pp. 71-80.

[2] Z. L. Lan, Z. M. Zheng, and Y. W. Li, "Toward automated anomaly identification in large-scale systems", IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 2, (2010), pp. 174-187.

[3] K. H. Ramah, H. Ayari, and F. Kamoun, "Traffic anomaly detection and characterization in the tunisian national university network," in NETWORKING 2006. Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems. Springer, 2006, pp. 136–147.

[4] C. Wang, K. Viswanathan, L. Choudur, V. Talwar, W. Satterfield, and K. Schwan, "Statistical techniques for online anomaly detection in data centers," in Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on. IEEE, 2011, pp. 385–392.

[5] C. Wang, V. Talwar, K. Schwan, and P.Ranganathan, "Online detection of utility cloud anomalies using metric distributions," in Network Operations and Management Symposium (NOMS), 2010 IEEE. IEEE, 2010, pp. 96–103.

[6] H. Qiu, N. Eklund, X. Hu, W. Yan, and N. Iyer, "Anomaly detection using data clustering and neural networks," in Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence).IEEE International Joint Conference on. IEEE, 2008, pp. 3627–3633.

[7] Husanbir S. Pannu, Jianguo Liu and Song Fu "AAD: Adaptive Anomaly Detection System for Cloud Computing Infrastructures", 31st International Symposium on Reliable Distributed Systems, pp. 396-397, 2012.