Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Machine Learning Approach For Data Security Audit In Cloud Computing

G P C Venkata Krishna[*1] & Dr D Vivekananda Reddy[2]

* 1 Research Scholar, Dept of Computer Science and Engineering, SVUCE, Sri Venkateswara University, Tirupati, Andhra Pradesh, India
*2 Professor, Dept of Computer Science and Engineering, SVUCE, Sri Venkateswara University, Tirupati, Andhra Pradesh, India

Ensuring the security of the cloud is one of the main barriers to wider cloud adoption. Potential cloud adopters want to know whether the con- trols in the cloud environment will adequately protect critical assets that are being migrated to the cloud.A surge of cloud computing has presented a number of challenges to the research community. Among them, identi- fying anomalous events is an immense challenge due to the complexity, heterogeneity and dynamic behavior of the event data. Anomalies and outliers are a significant security risk. Therefore, many types of research have been carried out on detecting anomalies in cloud data. Real time security auditing for the cloud plays an important role in mitigating major security concerns in the cloud. Most of the existing solutions are slow in response time and require lots of manual efforts. In this paper, we proposed a solution for Realtime machine learning based predictive security auditing for cloud which does two level (user/system) security audit at real time. The key to guaranteeing the cloud is near-real-time cloud auditing, which delivers immediate evaluation results and prompt response. We explore security and privacy risks in cloud computing, as well as the present state of cloud auditing initiatives, in this paper.

Keywords: *Cloud auditing, Machine learning, Cloud security, Outlier detection,Anamoly detection.*

1 Introduction

Attacks against cloud platforms have quickly increased in recent years. Cloud computing platforms, incidentally, accounting for more than 20% of all cyber attacks in 2020, making them the third most-targeted cyber environment. [1]Increased resource scalability necessitates dynamic compositions of comput-ing hardware, which introduces design flaws such as tenants sharing the same cloud and perhaps interacting by design. As a result, there are potentially hostile machines in the cloud network that must be guarded. Multiple levelsof computation security are breached by hostile machines on the network. A rival business may break into infrastructure components like hosts to obtain sensitive information about its consumers and other data that is kept on the system.[2]

Numerous anomaly detection technologies are being developed to find newassaults. Techniques for statistical detection and machine learning can be usedto find anomalies.[3]In the training phase of statistical detection and machine learning techniques, supervised anomaly detection algorithms are frequently employed in conjunction with labelled data or attack-free data to simulate normal behaviour. Attack-free training data is crucial for these algorithms' effectiveness. In a real-world network setting, it is challenging to find such training data.[4]

Unsupervised anomaly detection approaches include clustering and outlier identification in data security.They try to spot abnormal behaviour without using any understanding of the training set. The clustering technique DBSCAN-MP that we offer in this research utilises DBSCAN with multiple parameters to handle various traffic kinds.[5]

In this research, we present a proactive auditing system for securing severallevels of a cloud that is automated and efficient. To do this, we first offer an automated method for creating prediction models that capture the dependen-cies between cloud events. As a result, we are able to overcome the drawbacks of existing methods for manually capturing dependence relationships. Then,we use this model to forecast future occurrences so that our tool can auto- matically validate these expected events, which solves the problem of admins having to manually input future plans. Finally, at runtime, we simply checkthe pre-calculated verification findings to ensure that our method responds quickly, overcoming the inefficiencies of other approaches.[6]

1. This is the first time, to our knowledge, that a multi-level proactive secu- rity auditing solution has been proposed, which defends a cloud

against awide range of security threats directed at the cloud's various layers (i.e., user, virtual, and physical).

2. Unsupervised machine learning uses a model without a supervisor over-seeing it. By watching data and identifying patterns on its own, the data scientist lets the machine learn. In other words, this branch of machine learningenables a system to take decisions based solely on the information provided.

3. We connect our solution with Amazon EC2, a prominent cloud platform, and our experimental findings show that each request is responded to with littlelatency (e.g., a few milliseconds), demonstrating the efficiency of our system.

2 Related Work

2.1 EXISTING GAPS IN THE LITERATURE

The application potential of businesses has increased thanks to Cloud computing services. In other words, corporations can outsource excess elements of their IT infrastructure to Cloud computing, and the Cloud provider maybe responsible for system maintenance and servicing. This will minimize the financial load on businesses.[6] The most difficult task, however, is establishingtrust between the end customer and the cloud service provider. The end user may believe that he or she has no control over data stored elsewhere in the cloud data center. If this information is made available to end users, security techniques on the user's end may be used to protect the data.[6]

As we will see in the literature review, the majority of security-related issues concern stationary cloud data, whereas security for data in transit isnot addressed in the studied literature. In some situations, virtualization flaws can expose the cloud computing environment to a variety of security attacks and threats, including information leaking from committing data and service disruption owing to sharing and centralizing resources. Apart from that, if an attacker gains control of the hypervisor, there is no mechanism to control the attack. As previously stated, there are currently few approaches available to demonstrate where a user's data is stored in the cloud.[7]

The following section discusses a third-party auditing scheme that strength-ens the confidence between the end user and the cloud service provider by taking these issues into account. The security of the third-party audit-ing method is improved by using several types of available cryptographic approaches, which increases the trust factor between users and cloud service providers.[7] Depending on the type or extent of the audit, there are several dif-ferent types of cloud audits. An impartial team of auditors

normally conducts audits and looks at the potential of any cloud services offered. Internal audits are a less common choice because there may be prejudice in the analysis.

Methods used	pros	cons
1. Construction of Audit Internal Control Intelligent System Based on Blockchain and Cloud Storage	The blockchain storage application solution requires each node to keep all of the data, wasting storage resources while also failing to fully utilize network traffic routing resources. This research offers a unique AI and BC-based intelligent system for audit internal control based on this foundation. The simulation is used to design and validate the system. The proposed intelligent system has the potential to boost overall productivity.	1.Blockchain can slow down when there are too many users. 2.blockchain can not go back as this technology is immutable. 3.high cost
2.Auditing and Deduplicating Data in Cloud for Security	SecCloud and SecCloud+ are two secure technologies that we present. SecCloud combines an auditing entity with MapReduce cloud maintenance, allowing clients to establish data tags before uploading and check the integrity of data saved in the cloud. In comparison to prior efforts, the user computation in SecCloud throughout the file uploading and auditing phases is considerably decreased. SecCloud+ was created with the understanding that users will always want to encrypt their data before uploading it, and it allows for integrity auditing and secure deduplication on encrypted data.	1.Support for Batch Processing only 2.No Real-time Data Processing 3.Vulnerable by Nature
3.Privacy-Preserving Public Auditing for Shared Data in the Cloud	Using these current procedures to evaluate the integrity of shared data will inevitably divulge confidential information—identity privacy—to public verifiers. In this research, we present a unique privacy-preserving approach for public auditing of cloud-based shared data. We use ring signatures in particular to generate the verification information required to audit the validity of shared data. The identity	1. This method includes lots of mathematical calculations which are hard to understand and the level of complexity increases with the implementation. 2.Due to complex calculations this method is slower. 3.slow speed and costly.
4.A Tiered Strategy for Auditing in the Cloud	This paper presents a three-tiered method to cloud information auditing in this study. The method offers viewpoints on auditable events, such as the compositions of audit trails created independently. Stakeholders can identify potential security vulnerabilities and performance aspects by filtering and reasoning through audit trails.	1.Vulnerability to attack 2.Limited control and flexibility 3.Security and privacy
Third party public auditing on cloud storage using the cryptographic algorithm	It consists of three entities namely data owner, TPA and cloud server. TPA verifies the integrity of data on demand of the users. Thus no additional burden is provided on the cloud server. It is used only to save the encrypted blocks of data. All the tasks for the scheme are performed by the TPA and data owner. The introduced auditing scheme makes use of the AES algorithm for encryption, hash value to verify the integrity of the data and code regenerator when the data is corrupted. TPA(third party auditor)	

Fig. 1 Comparison of data security in cloud

2.2 Background on Cloud

Three levels of computer technology are used in cloud computing: SaaS (software as a service), PaaS (platform as a service), and IaaS (infrastructure as a service). Cloud computing evaluates user services utilising many layers of com-puter technology.[8] A cloud computing audit is comparable to other audits carried out within a company. Its major objective is to assess and enhance data

S.No	Paper	Approach
1.	**Toward a Real-Time Cloud Auditing Paradigm**	The amount of computing done in the cloud is greatly increasing. The decentralized nature of the cloud, however, makes it difficult for individuals to ensure that the computation is being done correctly. Thus, the concept of "cloud auditing" has appeared. As applications in the cloud become more sensitive, the need for auditing systems to provide rapid analysis and quick responses also increases.
2.	**A Security Problem in Cloud Auditing Protocols**	In 2013, subversion attack comes to publity again by Mikhail Bellare, who was inspired by PRISM. In this work, we implement this kind of attack on cloud auditing protocols. We show that through subversion attacks, the cloud server can recover the secret information stored by the data owner. Especially, First, we set a general frame of data auditing protocols.
3.	**A Machine Learning Auditing Model for Detection of Multi-Tenancy Issues Within Tenant Domain**	Cloud computing is intrinsically based on multi-tenancy, which enables a physical host to be shared amongst several tenants (customers). In this context, for several reasons, a cloud provider may overload the physical machine by hosting more tenants that it can adequately handle. In such a case, a tenant may experience application performance issues.
4.	**A Tiered Strategy for Auditing in the Cloud**	In this paper, we outline a tiered approach to auditing information in the cloud. The approach provides perspectives on auditable events that may include compositions of independently formed audit trails. Filtering and reasoning over the audit trails can manifest potential security vulnerabilities and performance attributes as desired by stakeholders.
5.	**Validating Cloud Infrastructure Changes by Cloud Audits**	One characteristic of a cloud computing infrastructure are their frequently changing virtual infrastructure. New Virtual Machines (VMs) get deployed, existing VMs migrate to a different host or network segment and VMs vanish since they get deleted by their user. Classic incidence monitoring mechanisms are not flexible enough to cope with cloud specific characteristics such as frequent infrastructure changes.

Fig. 2 Comparison of data in cloud - table2

accessibility while taking into account the general performance and security requirements that the cloud service provider should meet.[9]

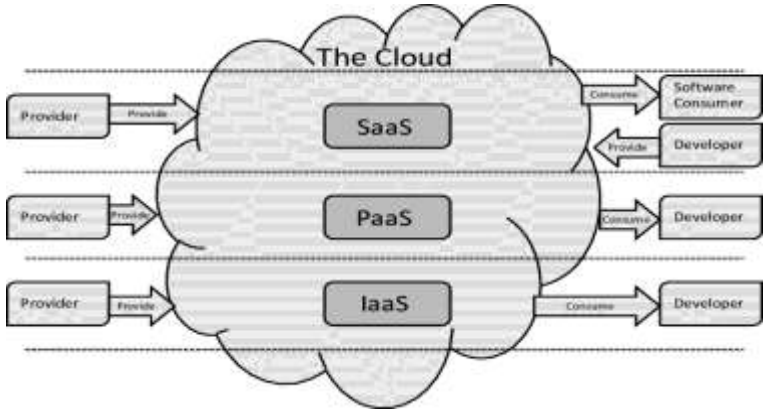

Fig. 3 Cloud Architecture

The privacy and security of any new technology are its most crucial components. The most crucial is the solution of the data integrity checking issue employing various systems security module. In the approach, the verifier's most crucial task can be divided into two categories: private auditability and public auditability. The data is remotely archived and updated by the client ina cloud computing architecture. The statistical data files seek attention over the data storage at far away locations, and the majority of this dynamically updated data has been obtained in limited time.[9]

In comparison to hard drives, USB drives, hard discs, computers, and other technologies, cloud computing has some disadvantages. Many government institutions are moving to cloud infrastructures, but worries about transferring sensitive data are growing.[10]

2.3 Amazon EC2

In our model we are using Amazon EC2, In order to assist us best fulfill the standards of our workload, Amazon Elastic Compute Cloud (Amazon EC2) offers the broadest and deepest compute platform, with over 500 instances and a choice of the newest processor, storage, networking, operating system, and purchase model. We are the first cloud with on-demand EC2 Mac instances,

400 Gbps Ethernet networking, and the only major cloud provider to support Intel, AMD, and Arm processors. We provide the lowest cost per inference instance in the cloud as well as the greatest pricing performance for machine learning training. AWS hosts more Windows, HPC, machine learning (ML), and SAP workloads than any other cloud.[11]

3 Preliminaries

3.1 Types of anomalies in the system

3.1.1 At user level

The administrative components of a cloud are included at the user level. This level includes a tenant's administrators and users, as well as their interactions with the cloud platform. Additionally, the cloud system's authentication and permission processes are incorporated in this level.[12]

3.1.2 Virtual level

The virtual level refers to the cloud's virtual components. This level especially considers virtual machines (VMs), virtual network components, and other vir- tual resources. This level also includes the procedures for managing thesevirtual resources as well as their relationships with other levels.[20]

4 Background on various Clustering Algorithms

4.1 What are Clustering Algorithms

Unsupervised machine learning tasks include clustering. Because of how this process operates, you could also hear it called cluster analysis. When using a clustering method, you will be providing the algorithm with a large amount of unlabeled input data and allowing it to identify whatever groups of data it can.These collections are known as clusters. A cluster is a collection of data points that are related to one another based on how they relate to other data points in the area. Pattern discovery and feature engineering are two applications of clustering.

4.2 Types of Clustering Algorithms

Different clustering techniques exist that can handle various sorts of unique data.

4.2.1 Density based

Density-Based According to the concept that a cluster in a data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density, clustering refers to unsupervised learning approaches that discover unique groups/clusters in the data.

4.2.2 Hierarchical based

An algorithm called hierarchical clustering, commonly referred to as hierarchical cluster analysis, divides objects into clusters based on how similar they are. The result is a collection of clusters, each of which differs from the others while having things that are generally similar to one another.

4.2.3 Centroid based

You probably hear the most about centroid-based clustering. Although it is quick and effective, it is a little sensitive to the first parameters you give it. These methods divide data points depending on several centroids present in the data. Based on its squared distance from the centroid, each data point is grouped into a cluster. This is the most widely utilised type of clustering.

4.2.4 K-means clustering algorithm

The most used clustering algorithm is K-means clustering. It is the most straightforward unsupervised learning approach and is centroid-based. This algorithm seeks to reduce data point variance inside a cluster. Additionally, it's how the majority of people first encounter unsupervised machine learning.Because K-means iterates through all of the data points, it is best applied to smaller data sets. As a result, if the data set contains a lot of data points, classifying them will take longer.

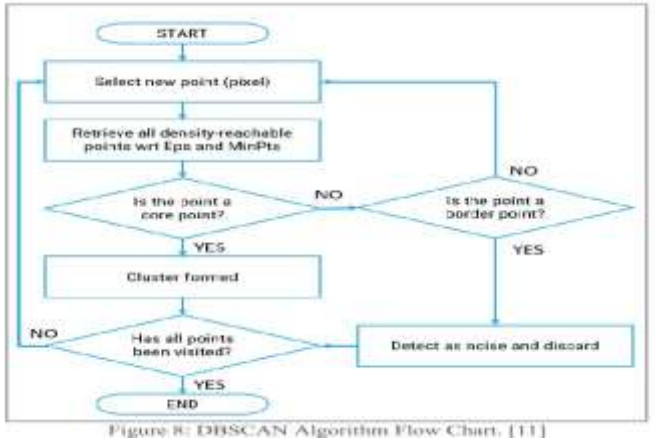

Figure 8: DBSCAN Algorithm Flow Chart. [11]

Fig. 4 workflow diagram

5 Methodology

DBSCAN: The Algorithm (In fig.3)

1. Lets assume an event P occurs at cloud
2. Retrieve all points density-reachable from p with respect to Eps and MinPts.
3. If p is a core point, a cluster is formed.
4. If p is not a core point, no points are density-reachable from p then checkif this point is a boarder point ignore it as its a noise. and DBSCAN visits the next incoming events of the cloud.
5. Continue the process until all of the points have been processed.

5.1 Research On K-MEANS CLUSTERING Algorithm and DBSCAN CLUSTERING Algorithm

5.1.1 Research on K-means Clustering Algorithm

Iterative clustering analysis is performed using the K-means clustering technique. The algorithm's main goal is to classify data based on their individual qualities without knowing the data's category or size. Prior knowledge is required for the K-means clustering process. The program divides the preprocessing data into k groups, chooses data from each group at random as the cluster center, and calculates the distance between each data in the group and the cluster center. Each object is assigned to the cluster center that is closest to it. A cluster is represented by the cluster center and the items assigned to it.The cluster center will cluster again based on the existing objects each timea sample is allocated. This procedure is repeated until the termination condi- tion is met. When the entire part of the item is good, no object is redistributed,and the cluster center is no longer altered, the classification is completed, and the classification error is minimized, this is the termination condition of the K-means clustering algorithm.

5.1.2 Research on DBSCAN Clustering Algorithm

The DBSCAN clustering technique is a density-based spatial data clustering approach that is the most often used. The clustering center is chosen by the algorithm as the region with the highest density. The approach is based onthe concept that any core object can decide clustering in a unique way. The

DBSCAN clustering technique uses the density clustering concept, which states that the number of objects in a given region of the clustering space must notbe fewer than the threshold. The Minkowski distance formula is a popular way of calculating similarity.

$$d\ is(X,\ Y) = \left(\sum^{n} |x_{i} - y|\right)^{\frac{1}{p}} \quad (1)$$

, the absolute distance is calculated, which is usually called Manhattan distance. When p=2, it expresses the Euclidean distance.

Any dense data collection can be clustered using the DBSCAN algorithm, which can also be used to quickly identify any data set outliers. It is frequently used to find anomalies in large data sets.

In the above calculation, when p=1

5.1.3 Why to choose BDSCAN not K-means Clustering

In K-means algorithm, K represents the number of clusters, and means represents the mean value of data objects in the clusters. K-means algorithm is a clustering algorithm based on division, which takes distance as the standard of similarity measurement between data objects, that is, the smaller the distance between data objects, the higher their similarity, they are more likely to be in the same cluster.

DBSCAN is a density based clustering algorithm. It defines a cluster as the maximum set of density connected points, which can divide the areas with enough high density into clusters, and can find clusters of any shape in the noisy spatial database. In this paper we use DBSCAN because we have to find noise point, and faster because there is no need to calculate any means value.

6 EXPERIMENTAL RESULTS

6.1 Experimental setup

The fundamental clustering algorithm DBSCAN is based on density based concept in which graph will be plot according to density of data. The fact that clusters of any shape can be detected is a benefit. The DBSCAN algorithm operates as follows: it examines each object's vicinity within the dataset. They are referred to as core objects if there are more objects in this area than MinPts. By accumulating locations that are easily reached from the core, each cluster develops from the core item. The method comes to an end if there are no more points that can be added to the cluster. These items are regarded as noise by the algorithm, unable to be grouped into any cluster.
Density-Based According to the concept that a cluster in data space is a contiguous region of high point density, separated from other similar clusters by contiguous regions of low point density, clustering refers to unsupervised learning approaches that discover unique groups/clusters in the data. The foundational algorithm for density-based clustering is called Density-Based Spatial Clustering of Applications with Noise (DBSCAN). From a big amount of data that contains noise and outliers, it may identify clusters of various sizes and forms.
There are two parameter of DBSCAN algorithm:-

minPts: The bare minimal quantity of points that must be grouped together a threshold for a location to be deemed dense.

eps (): A measurement of distance that will be used to find the points nearby any given point.

6.1.1 Data set

In this paper dataset is system generated random dataset is used, using that data we could plot the graph using matplot. Following the DBSCAN clustering,

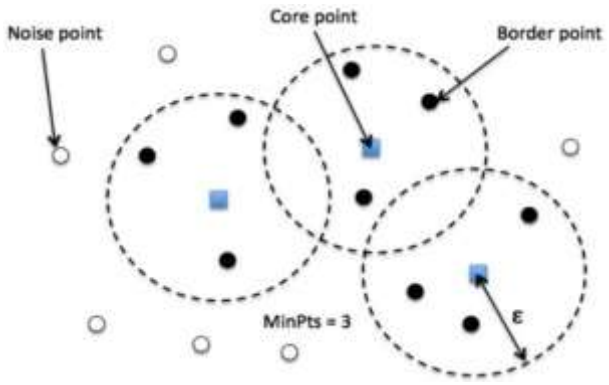

Fig. 5 Specification of points

there are three different kinds of points:
Core — This is a point that is n points away from at least m other points.
Border — This point has a minimum of one Core point at a distance of n.

Noise —This location falls outside of both the Core and the Border. Andit is separated from itself by fewer than m points within distance n.

6.1.2 DBSCAN clustering algorithm steps

1. The procedure starts by selecting a point in the dataset at random (until all points have been visited).
2. We consider all of these points to be a part of the same cluster if there are at least 'minPoint' points within a radius of epsilon to the point.
3. The neighbourhood calculation is then repeated recursively for each surrounding point to expand the clusters.

6.1.3 Parameter selection

The parameter problem exists in every data mining task. Each parameter has a unique impact on the algorithm. The variables epsilon and minPts are required for DBSCAN.

minPts: As a general rule, the number of dimensions D in the data set can be used to determine a minimal minPts, as minPts D + 1. The low value of minPts = 1 is illogical because every single point will then be a cluster on its own. The outcome of hierarchical clustering with the single - point metric anda cluster analysis cut at height will be the same when minPts 2. Therefore, minPts must be set to a minimum of 3. Larger values, however, typically perform better for data sets with noise and will produce more meaningful

	A	B	C	D
1	68.126832	161.6752	event1	
2	44.914873	75.058858	event2	
3	106.19347	142.42085	event3	
4	162.24187	120.27887	event4	
5	161.20529	119.60703	event5	
6	161.66116	116.45059	event6	
7	160.81865	120.43772	event7	
8	161.49271	118.91925		
9	164.75859	121.33459		
10	163.53175	122.37478		
11	159.58686	118.22116		
12	163.03078	116.18466		
13	157.20546	121.18857		
14	156.16201	118.28134		
15	157.90938	115.49903		
16	160.581	119.65721		
17	158.3429	118.01477		
18	159.44691	120.55184		
19	160.95882	122.2483		
20	160.36963	119.30043		
21	161.84871	117.76451		
22	162.08668	118.59442		
23	159.30307	122.2322		
24	162.7191	118.53772		
25	163.03479	116.38809		
26	164.50065	113.19134		
27	165.76116	117.09145		
28	162.84564	114.7918		
29	161.01057	113.32571		
30	166.88161	117.24191		
31	164.04552	115.53597		
32	159.46112	118.31983		

Fig. 6 co-ordinates of various cloud event

clusters. MinPts = 2 dim can be used as a general rule, but it could be essential to select bigger values for very large data, noisy data, or data that contains alot of duplicates.

epsilon :The distance to the nearest neighbour, k = minPts-1, ordered from the greatest to the smallest value, can then be plotted on a k-distance graph to determine the value for. If is selected excessively little, a significant portion ofthe data will not be clustered; nevertheless, if is chosen excessively high valueof , clusters will merge and the bulk of objects will be in the same cluster. Good values of are where this plot shows a ”elbow.” Small values of are generally preferred, and as a general rule, only a small portion of points should be close to one another.

Distance function: The results are significantly influenced by the selectionof the distance function, which is closely related to the selection of. In general, before choosing the parameter, it will be required to first determine a fair mea-sure of similarity for the data collection. This parameter cannot be estimated, but the relevant distance functions must be selected for the given data set.

6.2 Results

Real time input can be given by users to detect the anamoly in the system.Two types of events are given, user level events and system level events we can findthe anamoly in these events.

step1.User can view User level data and System level data.

step2.If user wants to change the existing data or update the existing data they can do the mandatory changes.

step3.Updated data will be saved, by clicking on audit cloud button theresult will be shown in form of cluster.

This is the snapshot given which we are using to find the output.

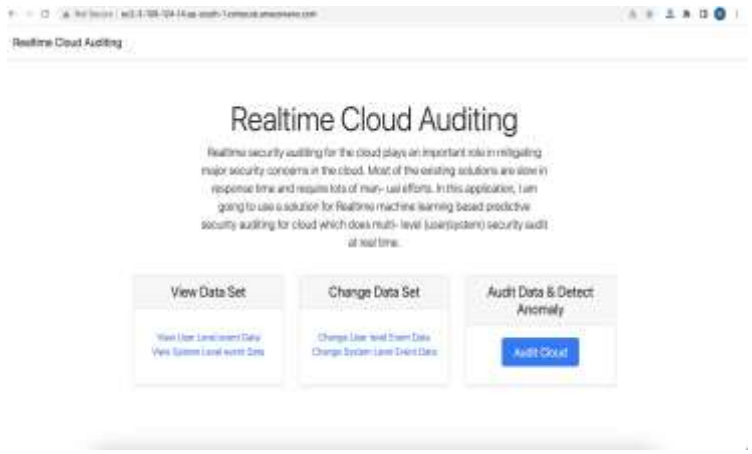

Fig. 7 User level events and systm level events

Clustering algorithm: choose the best clustering approach through exten- sive tests and theoretical analysis, then make the necessary adjustments tothe algorithm to enhance its performance in one or more areas for improved detection results. According to given cloud events(with co-ordinates)different clusters will be formed, clusters with high density considered as nonanamolousdata and clusters which are not following the parameters will be considered asnoise ponits.as in the following graph it shows.

7 Conclusion and Future work

The analysis of cloud security at the user and virtual level has been the primaryfocus of this paper. We also propose a real-time cloud security audit basedon the identification of potential anomaly events that may pose a threat to security. Without the use of a training dataset, our method, DBSACN, is appropriate for anomaly detection in cloud events in real time.

In order to find anomalies in time series data, we used the DBSCAN algorithm and contrasted it with a statistical anomaly detection method. The findings demonstrate that the DBSCAN algorithm can find anomalies even when they do not have extreme values.

We intend to use additional anomaly detection algorithms in the future to assess the algorithm's performance with various parameter configurations to identify potential security threats in real-time.

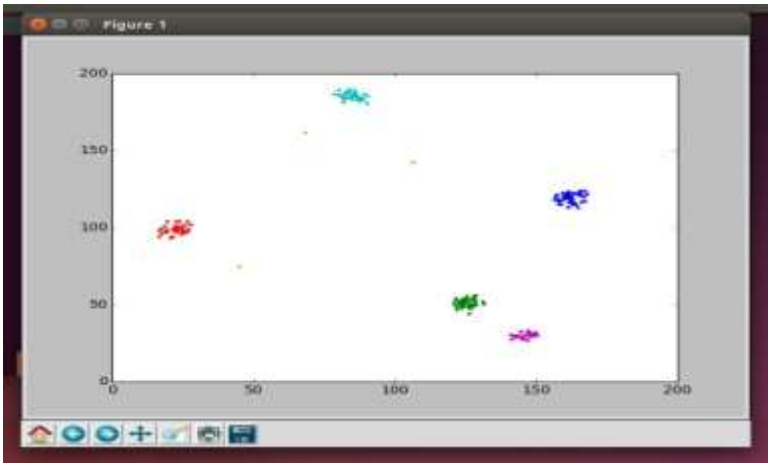

Fig. 8 clusters of the given data

References

[1] Peipei Zhou, Qinghai Ding, Haibo Luo et al., "Abnormal trajectory detection based on DBSCAN clustering algorithm", Infrared and Laser Engineering, pp. 230-237, 2017.

[2] Na Yin and Lin Zhang, "Application research of outlier mining based on hybrid clustering algorithm in anomaly detection", Computer Science, pp. 45-49, 2017.

[3] Jiakun Ruan, Yanguang Cai and Bing Le, "Freeway traffic flow abnormal data detection based on DBSCAN density clustering algorithm", Industrial Control Computer, pp. 75-79, 2019.

[4] S. Majumdar, Y. Jarraya, T. Madi, A. Alimohammadifar, M. Pourzandi, L. Wang, and M. Debbabi, "Proactive verification of security compliance for clouds through pre-computation: Application to OpenStack," in ESORICS, 2016.

[5] S. Majumdar, Y. Jarraya, M. Oqaily, A. Alimohammadifar, M. Pourzandi,

L. Wang, and M. Debbabi, "LeaPS: Learning-based proactive security auditing for clouds," in ESORICS, 2017.

[6] Y. Luo, W. Luo, T. Puyang, Q. Shen, A. Ruan, and Z. Wu, "OpenStack security modules: A least-invasive access control framework for the cloud,"in CLOUD, 2016.

[7] T. Madi, S. Majumdar, Y. Wang, Y. Jarraya, M. Pourzandi, and L. Wang, "Auditing security compliance of the virtualized infrastructure in the cloud: Application to openstack," in CODASPY. ACM, 2016, pp. 195– 206.

[8] Yingjie Yan, Gezhen Sheng, Yadong Liu et al., "Transformer state anomaly detection based on sliding window and clustering algorithm", High voltage technology, pp. 4020-4025, 2019.

[9] Xiaoqing Yu and Linhai Qi, "Power big data anomaly detection based on stream data clustering algorithm", Electric Power Information Technology, pp. 8-14, 2020.

[10] Pu Wang, Yusha Xiong and Chengcheng Wang, "Traffic anomaly detec- tion method based on path travel time analysis", Journal of University of Electronic Science and Technology of China, pp. 71-77, 2018.

[11] S. Majumdar, T. Madi, Y. Wang, Y. Jarraya, M. Pourzandi, L. Wang, and M. Debbabi, "Security compliance auditing of identity and access manage- ment in the cloud: application to OpenStack," in CloudCom. IEEE, 2015, pp. 58–65.

[12] Dandan Zhang, Ziyi You and Jian Zheng, "Optimized clustering algorithm based on improved local anomaly factor detection", Microelectronics and Computers, pp. 49-54, 2019.

[13] S S Li, "An Improved DBSCAN Algorithm Based on the Neighbor Similarity and Fast Nearest Neighbor Query", IEEE Access, pp. 99, 2020.

[14] Zhang Wang and Liang, "Short-Term Wind Power Prediction Using GA- BP Neural Network Based on DBSCAN Algorithm Outlier Identification", Processes, vol. 8, no. 2, pp. 157, 2020.

[15] Ming Zhao, Hongju Yan, Mingjun Zhang et al., "A network anomaly detection method based on clustering algorithm", Computer and network, pp. 68-71, 2020.

[16] Pu Wang, Yusha Xiong and Chengcheng Wang, "Traffic anomaly detec- tion method based on path travel time analysis", Journal of University of Electronic Science and Technology of China, pp. 71-77, 2018.

[17] Yasser El-Sonbaty, M. A. Ismail, Mohamed Farouk ”An Efficient Density Based Clustering Algorithm for Large Databases”, IEEE, ICTAI 2004.

[18] Glory H. Shah, ”An Improved DBSCAN, ”A Density Based Cluster-ing Algorithm with Parameter Selection for High Dimensional Data Sets”IEEE 2013.

[19] H He and Y Tan, ”Automatic pattern recognition of ECGsignals using entropy-based adaptive dimension ality reduction and clustering[J]”, Applied Soft Computing, vol. 55, no. 1, pp. 238-252, 2017.

[20] C Peng, Z Kang, F Xu et al., ”Image Projection Ridge Regression for Subspace Clustering[J]”, IEEE Signal Processing Letters, vol. 24, no. 7,pp. 991-995, 2017.