ISSN:0975-9646

Secure Data Access control in Cloud Environment

[1] G. Praveen Babu, [2] B. Sushma Rao

[1] *School of Information Technology, JNTU, Hyderabad, India*

[2] *M.Tech, S.I.T, J NTU, Hyderabad, India*

Abstract **- Cloud computing is most prevalent technology used in various organizations. It provides wide range of applications that are useful for various business needs. It provides the facility to store and access the data from any place around the world across distributed network environment. Data stored in the network is always vulnerable to threat from an outsider and as well as from a malicious insider. Various encryption mechanisms have been used to provide data security. But they have failed to provide enough protection to the data.**

We propose a different approach for providing security to the data in the cloud by using deception technique. We monitor the different user's search behaviour to analyze their search patterns. Using these different patterns we detect abnormal data access patterns. Once illegal data access is identified and verified by answering security question, we launch disinformation attack and ensure to minimize data theft and misuse of user's real data.

Keywords: **Data access, Data security, malicious insider, decoy information, SVM, multi clouds, big data analytics.**

I. INTRODUCTION

Cloud computing provides users with flexible, Cost Efficient, limited Storage, Backup, Recovery, Easy Access of Information and provides better Deployment facilities to the user's of the cloud. Businesses like small, medium, and large companies prefer to outsource their services to cloud for better operational efficiency. Cloud is just not a commodity plan for cutting the cost of running a few applications, but it is also about transforming business.

Cloud has 3 service models IaaS, PaaS, SaaS. The bottom layer is Infrastructure as a Service (IaaS) provides virtual machines and other resources like block and files storage, network Security, load balancing, virtual local area networks (VLANs) etc. IaaS cloud providers supply these resources on-demand from their large pools installed in their centres. For wide-area connectivity, customers can use either the Internet or carrier clouds (i.e. dedicated virtual private networks).

The second layer from the bottom is Platform as a Service (PaaS) here, cloud service providers deliver a computing platform like operating system, execution environment (programming language), database, and web servers. Application developers can develop and run their software on a cloud platform without buying and managing the hardware and software layers. Some PaaS service providers like Windows Azure, Google AppEngine enable the computers and storage resources vary automatically to match application demand so that the cloud user does not have to allocate resources manually.

The last service model is Software as a Service (SaaS), users are provided access to application software and databases. Cloud service providers manage the infrastructure and platforms that run the applications. SaaS is sometimes also called as "on-demand service of software" and is usually available on a pay-per-use basis. Cloud service providers install and operate application software in the cloud and cloud users can access the software from cloud clients. Cloud users need not manage the cloud infrastructure and platform where application runs. This reduces the need to install and run the applications on the cloud user's own computers. This simplifies maintenance and support.

E.g. Salesforce, SalesForce CRM, Google Apps, DeskAway, Impel CRM, Wipro w-SaaS are some of the SaaS examples.

II. NEED OF SECURITY IN CLOUD

Outsourcing is advantages because of its reduced overhead, low operational control, staffing flexibility and risk sharing. Sharing data in cloud is associated with great risk and most common known risk is data theft.[1] Some of the important security issues related to cloud are:-

1) Network Threats and Attacks: Using cloud-based services requires organizations to allow their networks to connect with multiple cloud service providers networks. This increases the risk to negotiate on one network to gain remote access to other network. The user need to able to make these connections securely without effecting new options.

2) Data Protection and Access: The cloud service provider owns the infrastructure, platform and service. But, the user still owns the data. When employees, partners, users and administrators connect to SaaS, cloud service need to validate that data will be protected from any other tenants of cloud environment.

3) Compliance: Cloud vendors can monitor regulatory compliance and security governance rules, to know about how are the user identity and access handled? How are the administrators access monitored? How the user activity is monitored, anomalies detection and incident reporting handled? We have security policy and governance model for IT department and, one should not compromise that when we move to the cloud. These rules can help the user to feel that his data will be secured from illegal access.

2.2 Data Protection and Access

Data security is one of the important areas which needs importance. Data theft can happen either by an insider or an outsider. An attack by an outsider is most common type of attack. User's private data such as bank account details, online purchase details, bank receipts and any other important data are more prone to attack. User's credentials are hacked without his knowledge and misuse of the data can take place. To stop such type of attacks, there are many encryption algorithms like RSA, DES IDEA, Digital signatures, AES, Homomorphic encryption etc that provide better security to the data.

But finding a malicious insider [2] using these algorithms do not guarantee to provide security[3] because authentication and authorization is given to every valid registered members of the cloud.

Intruder detection using log analysis was studied by [4] but we try proposing a different approach by analyzing search patterns of different user's of the cloud.

III. SECURING THE CLOUD

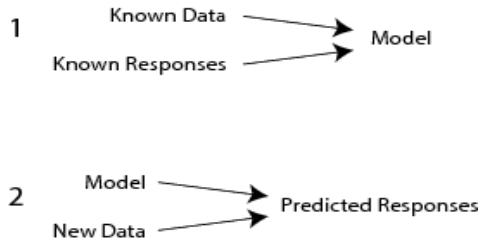
3.1 Data Protection and Access

Machine learning algorithms have gained lot of attention in field of computer science .machine learning is the sub branch of artificial intelligence, that build up and studies the systems that can learn from data. For example, a machine learning system can be used trained on email messages that can learn to identify difference between spam and non-spam messages. After learning from known data, the model can then be used to classify new email messages into spam and non-spam folders.

Machine learning uses prediction on the data, based on known properties, and learns from the training data. Machine learning has wide range of algorithms which are divided into 2 types supervised and unsupervised learning.

3.2 Supervised learning

Supervised learning machine learning [5][6] takes known set of input data and known responses to the data, and seeks to build a predictor model that generates reasonable predictions for the response to new data.



Support Vector Machine algorithm [5] [6] is the most popular algorithms in supervised learning methods, in the modern machine learning. It was introduced by Vapnik in 1992 which is based on statistical learning and provides better classification than other machine learning algorithms. It is computationally very expensive because it involves data matrix inversion. Few examples of SVM applications include text categorization, image segmentation, bioinformatics and Hand written character recognition. SVM is applied to the data points as mentioned in the diagram

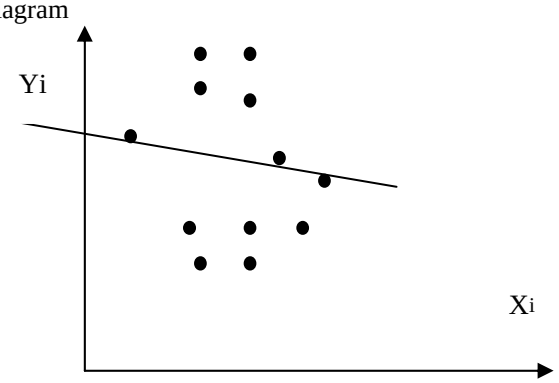


Fig.1 data points with different margin

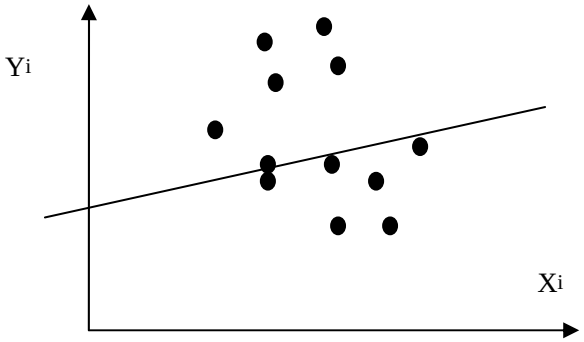


Fig.2 – data points with different margin

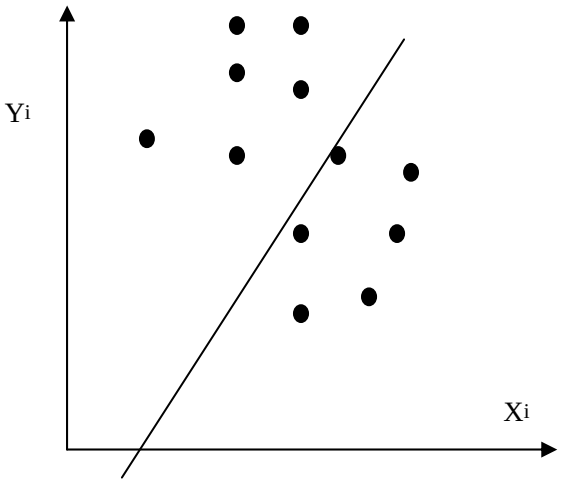


Fig.3 data points with another alignment

The above mentioned figures suggest that selecting an optimal margin to separate the data points is difficult. The points that lie on the line or near the line cannot be placed on either side of the separator. The points that lie on either side of the line are divided into 2 classes i.e. +1 & -1. Data points in each class that lie closet to the margin or classification line are called as "support vectors", after training we can discard all the data except for support vectors and use them for classification.

We use straight line equation **y = w . x + b, w** is the weight vector, **x** is the input vector, **b** is the bias weight.
The classifier line or margin is any value of **x** that gives positive above the line
i.e. **w . x + b > 0** the point is above the line which is +1 class and
w . x + b < 0 the point is below the line which is --1 class

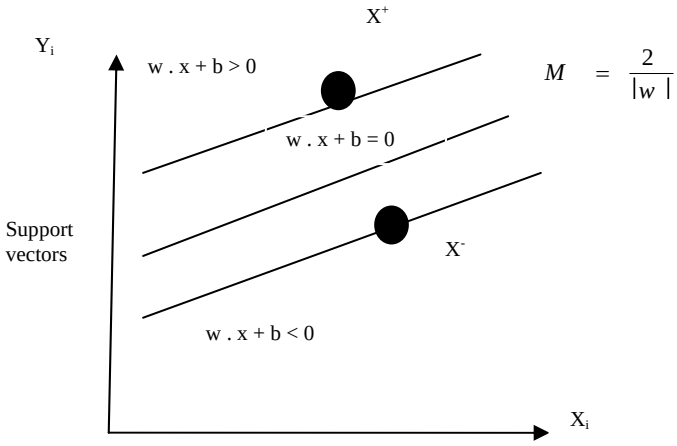


Fig.4 data points with another alignment

The margin $2 / |w|$ is separator distance between support vectors.

3.3 disinformation attack

A user's file system is different from every other user's of the cloud. Because the user knows his file system, he knows where his data is stored, but malicious insider is unaware of the user's file system and therefore searches for the data. Here we can profile [8] difference of search pattern from valid user to an intruder. Once we have finished indentifying malicious insider, system asks the user's of the cloud to enter security key (challenge question) which is known only to the valid user. Intruder obviously enters wrong key for known reasons. This is where we launch disinformation attack, which is assumed to be correct by the intruder. i.e. decoy information. [9]

In this way we can stop unauthorized access of the data and minimize data theft in the cloud environment. All unsuccessful attempts are maintained in the log file for further analyses. We identify difference between normal files and decoy files by attaching HMAC to the user's file for maintain authenticity of the data. Thus SVM ensures to minimize data theft in the cloud when compared to other machine learning algorithms[7]. SVM has limitation on the size of the data sets, It cannot be used on large data sets.

IV. EXTENDING SVM TO MULTI CLOUDS

Intercloud or multi cloud is interconnection of network of networks i.e. two or more number of clouds connected across the network. The Intercloud concept is based on the key factor that each single cloud has limited physical resources. If a cloud saturates its computational and storage resources as part of infrastructure, or is requested to use resources which cannot be extended beyond its capacity, it will still be able satisfy such requests for service allocations sent from clients. The Intercloud theory would address situations, where each cloud will use computational, storage, or other resource of the infrastructures of other clouds. It is similar to mobile technology where operators implement roaming and interoperability to their mobile customers.

Exchange of data in cloud, peering, and roaming introduce new business opportunities among cloud service providers. When there is a need for data transfer in intercloud's environment, there has to be a provision for data security. Integrity, confidentiality, authentication should be maintained in multi clouds environment. [11] Cloud service providers should take adequate measures for reliable data transfer and data access. Using SVM in multi clouds environment helps to reduce data theft. In multi clouds large data sets [10] needs to be processed, we can use BIG DATA Analytics technique called HADOOP. HADOOP [12] is a product Apache software foundation. It is an open source framework which gained much popularity in recent times. Map and Reduce functions are the two most common techniques in HADOOP that can help in processing large data sets. These functions can be used along with SVM to avoid unauthorized data access.

V. CONCLUSION

In this paper, we present a novel approach for securing personal and business data in the Cloud. We also propose ways to monitor data access patterns by profiling user behavior to determine abnormal data access. Once unauthorized data access or exposure is suspected and later verified, with security questions, we can inundate the malicious insider with decoy information and minimize data theft. Decoy documents are stored in the Cloud along with user's real data serve as sensors to detect illegitimate access. Preventive attacks that rely on disinformation technology can provide quite a good level of security in the Cloud.

REFERENCES:

[1] Akhil Behl, kanika Behl "An Analysis of cloud computing Security issues" 2012 world congress of information and communication technology.

[2] Lucky Nkosi, Paul Tarwireyi and Matthew O Adigun "Insider Threat Detection Model for the Cloud", 978-1-4799-0808-0/13/$31.00 ©2013 IEEE.

[3] M. Van Dijk and A. Juels, "On the impossibility of cryptography alone for privacy-preserving cloud computing," in Proceedings of the 5th USENIX conference on Hot topics in security, ser. HotSec'10. Berkeley, CA, USA: USENIX Association, 2010, pp.1–8.[Online].Available:http://dl.acm.org/citation.cfm?id=1924931.1924934

[4] Manish Kumar, Dr. M. Hanumanthappa, "Scalable Intrusion Detection Systems Log Analysis using Cloud Computing Infrastructure", 978-1-4799-1597-2/13/$31.00 ©2013 IEEE.

[5] Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995.

[6] Wencai Zeng, Jiong Jiay, Zhonglong Zhengz, Chenmao Xie and Li Guo, "A Comparison Study: Support Vector Machines for Binary Classification in Machine Learning", 2011 4th International Conference on Biomedical Engineering and Informatics (BMEI).

[7] Lucky Nkosi, Paul Tarwireyi and Matthew O Adigun "Detecting a Malicious Insider in the Cloud Environment Using Sequential Rule Mining", 978-1-4799-8/13/$31.00 ©2013 IEEE.

[8] M. Ben-Salem and S. J. Stolfo, "Modeling user search-behavior for masquerade detection," in Proceedings of the 14th International Sympo- sium on Recent Advances in Intrusion Detection. Heidelberg: Springer, September 2011, pp. 1–20.

[9] B. M. Bowen and S. Hershkop, "Decoy Document Distributor:http://sneakers.cs.columbia.edu/ids/fog/,"2009.[Online]. Available:http://sneakers.cs.columbia.edu/ids/FOG/

[10] Xiaoou Li, Jair Cervantes, and Wen Yu, "Two-Stage SVM Classification for Large Data Sets via Randomly Reducing and Recovering Training Data", 1-4244-0991-8/07/$25.00/©2007 IEEE

[11] Mohammed A. AlZain , Eric Pardede , Ben Soh , James A. Thom, "Cloud Computing Security: From Single to Multi-Clouds", 2012 45th Hawaii International Conference on System Sciences

[12] Kiran M., Saikat Mukherjee, Ravi Prakash G. "Verification and Validation of Parallel Support Vector Machine Algorithm based on MapReduce Program Model on Hadoop Cluster", International Journal of Computer Science and Management Research eETECME October 2013,ISSN 2278-733X