Design and Implementation of Secure Multi-Authentication Data Storage in Cloud using Machine Learning Data Classification

Amanpreet Singh
Assistant Professor
Department of CSE

Manju Bala, PhD
Associate Professor
HOD of CSE Dept.

Supreet Kaur

ABSTRACT
Cloud computing offers numerous benefits including scalability, availability and many services. But with its wide acceptance all over the globe, new risks and vulnerabilities have appeared too. Cloud computing supplies facility of storing and accessing understanding and programs over the web without bothering the storage space on procedure. Storing the data on cloud eliminates one's worries about space considerations, buying new storage equipment or managing their data, rather they are able to access their data any time from any place provided they have internet access. However, the rising security issues have resisted the companies from connecting with cloud computing fully. Hence security risks have appeared as the main disadvantage of cloud computing. This paper involves the efforts to research the security risk and then proposes a framework to address these risk on the authentication and storage level in cloud computing. While addressing the security issues the first and the foremost thing is to classify what data needs security and what data needn't bother with security and hence data gets classified into two classes sensitive and non-sensitive. To achieve data classification, a data classification approach based on the confidentiality of data is proposed in this paper. Following that an efficient security mechanism must be deployed by means of encryption, authentication, and authorization or by some enhanced means of security techniques to ensure the privacy of data on cloud storage.

Keywords
Cloud Computing, Data confidentiality, Security, Data Hiding, Machine learning, Cloudsim, Datacenter, Resource, Data Privacy

1. INTRODUCTION
Arising strategy nowadays is cloud computing. As of late it is discovered that investigators have interest in utilizing cloud for carrying out technical applications and also the enormous associations are on the edge of changing over to hybrid cloud. Numerous applications which are very complex need parallel processing for executing the jobs efficiently. Because of the synchronization and communication among processes which run parallel, there is a reduction in usage of resources of CPU. It is fundamental for a data center to accomplish the use of hubs while keeping up the level of responsiveness of jobs which are running parallel. The cloud computing is pulling in an expanded number of uses to keep running in the data centers which are remote. Numerous intricate applications necessitate capabilities of parallel processing. A portion of the applications which are running parallel demonstrate a decline in usage of resources of CPU at whatever point there is an expansion in parallelism only when there is no planning of jobs accurately then it lessens the execution of a computer. The offered Cloud Service Models classified as Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS).

Cloud Computing is not completely perfect, because there are many aspects that are needed to be worked out like security issues. The cloud model still suffers from significant security challenges. The consumers are raising questions like where their data are hosted, non-standard authentication methods, exposure of data to foreign entities. These questions are slowing down the adoption of Cloud Computing technologies [1]. Data Loss, Service Traffic Hijacking, Insecure Interfaces and APIs, Denial of service, Malicious Insider and Cloud abuse are some of the cloud security issues being faced nowadays.

Application level security issues include Risks at the application level that directly affect the cloud applications at a user level. Solutions supplied to lessen these kinds of issues are the encryption techniques and DDoS detection methods to prevent DOS attacks. Secondly, Network level security issues involves attacks like DNS poisoning attacks, phishing attacks, IP spoofing attack. However, these attacks can be carried out either internally or externally. To address these problems efficient servers use security ensures trust configuration uses SSL and IDS and lastly, data storage level security issue also includes various challenges like Data-in-Transit, data loss, Data integrity and Data stealing. [2]

This research work concentrates on privacy issue in cloud computing. Whenever, the information is exchanged to the cloud server it experiences a security system i.e. encryption without comprehension the level of sensitivity of the data or the data is essentially put away on cloud server without securing it. All information has diverse sensitivity levels so it is improper to store the information without comprehension its sensitivity level and security necessities. So, Basically, Data Classification is being presented over here in this work. To direct the security requirements of data, a data classification model has been proposed to classify the data according to its sensitivity level and then encrypting the only data which is required to secure using an encryption technique in cloud environment. A very intelligent technique to secure the data would be to first classify the data into sensitive and non-sensitive data and then secure the sensitive data only. Efficient security mechanisms should be deployed by means of encryption, authentication, and authorization or by some other method to ensure the privacy of consumer's data on cloud storage.

2. RELATED WORK
Deyan Chen et.al [3] analysed the problems of data security and privacy protection. Many organizations today are

realizing benefits by making their applications and data in the cloud. Adoption of cloud computing can lead to improvements in efficiency and effectiveness in the development and deployment and save the cost. Data security issues and the protection of privacy remains the main inhibitor to the adoption of cloud computing services. This paper gave a concise but all round analysis on issues of data security and privacy protection associated with cloud computing through all stages of the lifecycle of the data. Then, this article discussed some of the current solutions.

Guo M.-H. M, Liaw[4]worked on graphical passwords for authentication system. In that work, they used two main functions which are cloud devices and cloud environment. Here, main purpose was to authenticate users from authentication server. Here, the user using their cloud device first run cloud and then enter their userID and the program shows has some graphs from where user selects a point on the graph. Now program generates public and private key which applies when the message is transmitted. Now, when the message is generated by user onto cloud new time stamp is added in the transmission. Now each time when message is retrieved from the cloud, the hash values are used to validate the message.

Prathamey K. Rane et.al [5] described all graphical methods for password authentication system and also proposed an approach which described that first calculation has been done by server based on user entered username and according to result one set of images will be transferred on user screen, each set contains hundreds of images, and then user has to select two images from given set, whereas server also add two images by its own to form complete password.

Mohammed Faez Al-Jaberi et.al [6]proposed an architecture based model that provided data integrity verification and privacy preserving in cloud computing. This model has presented an effective mechanism that provides data integrity verification without allowing third party to violate the privacy of data. AES and MD5 algorithms are used for data integrity and privacy is ensured against unauthorized parties.

Raghul Mukundan et.al [7] proposed that a cloud service provider (CSP) is boon for data owners, to outsourcing their data and reduce their burden of local data storage and maintenance. Cloud service provider replicates the data to increase the data availability, reliability and durability. And clients or data owners have to pay to store data to CSPs storage place. To maintain the data confidentiality stops the cloud service provider from cheating by maintaining fewer copies than paid for data, in this paper they propose the Dynamic Multi Replica Provable Data Possession scheme (DMR-PDP). This scheme also provided service of dynamic operations such that insertion, deletion and modification on replicated data over the cloud server.

An Na Kang et.al [8] described that the cloud computing is most widely used service that provides a variety of computing resources, from servers and storage to enterprise applications like email, security and backup all delivered over the internet. Cloud computing also manages the user's IT resources as well as enterprises the IT resources in an effective manner. With the development of internet cloud computing effectively manage and use the number of data. While using cloud computing so many threats have occurred, since they give rise to security threats to enterprise information. Authentication and access authority management was the basic idea to be considered to protect information leakage, service abstraction etc. Core factors that cloud computing use to protect data are management of keys and powerful encoding.

Ruhui Zhang et.al [9] described that security of cloud computing is main concern now a day's. In this work virtualization was used for security purpose of cloud computing that make safe the cloud. In which virtual machine monitor (VMM) scans the infected data. So firstly, it collects behaviour of the process and calculates the distance to check whether malevolent behaviour occurs, this is calculated through intrusion detection method.

Deyan Chen et.al [10] analysed the problems of data security and privacy protection. Many organizations today are realizing benefits by making their applications and data in the cloud. Adoption of cloud computing can lead to improvements in efficiency and effectiveness in the development and deployment and save the cost. Data security issues and the protection of privacy remains the main inhibitor to the adoption of cloud computing services. This paper gave a concise but all round analysis on issues of data security and privacy protection associated with cloud computing through all stages of the lifecycle of the data. Then, this article discussed some of the current solutions

Abuhussein A et.al [11] described the movement of client data to the cloud provider. However, for new customers choosing the right insurance services is sometimes difficult. So, in this article, the author has described some of the attributes that identify the policies of security and privacy services. These attributes helped for new customers become well known for the various services and their privacy policies. These security attributes are the backup, encryption, authentication and access control, hardware and isolation of data, data storage locations, monitoring, SLA compliance ,disaster recovery, protection of client-side and many more. The author attempted to identify and categorize a list of attributes which reflect the various aspects of cloud security and privacy. These attributes can be further used to access and compare the cloud computing services so that users can get the better cloud solutions.

Almorsy M, Grundy J, and Ibrahim A. S [12] demonstrated on the framework that described the FISMA rule allowing the agreement of trust between service consumers and service provider. Author described that due to the lack of restrictions on a service level agreement culminated in the loss of trust in cloud services. But this problem can be minimized if we are using standards such as FISMA and NIST-ISO 27000. To manage the security, process this framework enables collaboration between service providers and consumers to use the standard FISMA best suited to the cloud specification security and meet consumer need. The main work in this paper was dependent upon the security categorization of the service.

Singh Amritpal and Singh H [13] proposed an enhanced LSB based Steganography procedure for images bestowing better data security. It exhibits an embedding algorithm for hiding ciphered messages in nonadjacent and irregular pixel areas in edges and smooth regions of images. The edges in the cover-image are detected using improved edge detection filter. The encrypted message bits are then embedded in the least significant byte of randomly selected edge pixels and some specific LSBs of red, green, blue components respectively. Such type of steganography technique ensures least chances of suspicion about message bits hidden in the image and it gets hard to estimate the true message length by standard steganography detection methods. The Proposed approach

showed better results in PSNR value and Capacity as compared to other existing techniques.

Hamid Banirostam et.al [14] described one new approach named Trusted Cloud Computing Infrastructure (TCCI) which was based on Infrastructure security. TCCI approach describes that different nodes are required to run on secure environment so to keep hackers away. Moreover, if node runs in a secure environment than even administrator is incapable of access the user data. To make the infrastructure secure TCCI approach is proposed which handles the nodes by third party known as Trusted Coordinator (TC).

3. PROPOSED METHODOLOGY

The research involves exploring various security issues in cloud environment at with respect to three aspects authentication, confidentiality, integrity and analyzes their impacts. Also, it explores various data classification algorithms in machine learning like KNN, Naïve Bayes and Ensemble Learning and analyzes their performance.

The paper proposes a secure data classification model using Modified Ensemble learning technique supervised machine learning approach. In this, data is classified according to its sensitivity level. Then encrypting only, the data which is required to be secure using randomized and anonymized privacy preserving techniques embedded in steganography technique in cloud environment. [10] The proposed work also ensures the security by using single cloud provider and dividing single cloud into different zones thereby saving a cost of the client and also enhancing the security.

Step 1: Authentication Level:

a) **Owner-** top level security like giving the access after finger print scan, various security questions.

b) **Administrator-**second level security i.e. after asking various security questions then provides the access.

c) **User-** third level security i.e. providing access after username and graphical password matching.

This password is based on the sequences of some images. It is much secure because sequence of images is change every time. Basically, this password is use for authentication purpose. Only legitimate user will allow entering in cloud, if they enter the correct sequence of image. After authentication, during access of data operations this interface will again ask the user sequence, this time images gets shuffle, based on sequence of images password will also be change.

Step 2: Data Classification: Classifying the dataset by using the Modified Ensemble learning technique.

1. Decision tree is used as Meta classifier.

2. Meta Learner is a learner scheme that combines the output of the Improved Ada boost and bagged J48 i.e. the base learners. The base learners' level-0 models and the meta-learner is a level-1 model. The predictions of the base learners are input to the meta-learner.

This will classify the data into: basic, confidential and highly confidential using the rules induced in the learning algorithms which will identify which attributes of the data set are under vulnerability attacks.

Step3: Data hiding Architecture: To keep the highly confidential data secure from attackers on the network, data is hidden inside the image using randomized and anonymized privacy preserving techniques embedded in steganography technique. Then sending that image data file to the cloud environment. Inference channel analysis is a control used in the output of dataset analysis in order to stop a person who has access to only summary information from being able to determine (infer) a particular value for a particular record. The severity estimation of the datasets is identified in order to perform inference channel analysis to avoid privacy leakage

Step 4: Enhancing the security by using single cloud provider and dividing single cloud into different zones thereby saving a cost of the client and also enhancing the security. Then segregation of data will be done by creating virtual partitions of data for saving and allowing user to access data in his partition only. Each user will have the rights according to the role of the client i.e. role based access polices Use of virtual partition and enhanced user access control in cloud system will improve data security and thereby fixing the threats in data mining to Personal/private data in cloud systems. Enhanced Cloud system will be compared with existing secure cloud systems. We will compare enhanced system against security, performance & ease of use.

4. PERFORMANCE PARAMETERS

The evaluation parameters considered for evaluating the performance of the proposed system are:

a. Classification Accuracy

b. Precision

c. Recall

d. Data Encryption time

e. Data Decryption time

5. CONCLUSION AND FUTURE SCOPE

In this research, a technique for secure authentication and data confidentiality in cloud environment is proposed. The focus of the research was to characterize the data taking into account the security prerequisites of the information that divides the data into sensitive and non-sensitive using modified machine learning algorithm. The fundamental contribution of this security model is data confidentiality and classification of data using machine learning classification approach. The classified confidential information is then encrypted using privacy preserving approach and is stored in the cloud server a while the non-confidential data is sent to the cloud environment as public data directly. Furthermore, to enhance the security at the authentication level, multi-level authentication mechanism is used including image sequencing passwords based on different themes has been used in order to avoid un-authorized access to the cloud environment. The proposed system has been simulated in a designed cloud simulation environment using cloud-sim simulator.

6. ACKNOWLEDGMENT

The paper has been composed with the kind assistance, guidance and support of my department who have helped me in this work. I would like to thank all the people whose encouragement and support has made the fulfillment of this work conceivable.

7. REFERENCES

[1] H. Tianfield, "Security issues in cloud computing," 2012 IEEE Int. Conf. Syst. Man, Cybern., pp. 1082–1089, 2012.

[2] Almorsy, M., Grundy, J., & Ibrahim, A. S. (2011) "Collaboration- Based Cloud Computing Security Management Framework" IEEE conference of cloud computing, Washington (DC), pp. 364-371

[3] D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012), vol. 1, pp. 647–651, 2012.

[4] M.-H. M. Guo, H.-T. H. Liaw, L.-L. Hsiao, C.-Y. Huang, and C.-T. Yen, "Authentication using graphical password in cloud," 2012 15th International Symposium on Wireless Personal Multimedia Communications (WPMC), pp. 177–181, 2012.

[5] S. M. Gurav, L. S. Gawade, P. K. Rane, and N. R. Khochare, "Graphical password authentication: Cloud securing scheme," 2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies, pp. 479–483, 2014.

[6] M. F. Al-jaberi and A. Zainal, "Data Integrity and Privacy Model in Cloud Computing,"2014 International Symposium on Biometrics and Security Technologies (ISBAST), pp. 280–284, 2014.

[7] R. Mukundan, S. Madria, and M. Linderman, "Efficient integrity verification of replicated data in cloud using homomorphic encryption," Distributed and Parallel Databases, vol. 32, pp. 507–534, 2014.

[8] A. N. Kang, L. Barolli, J. H. Park, and Y.-S. Jeong, "A strengthening plan for enterprise information security based on cloud computing," Cluster Computing, pp. 1–8, 2013.

[9] R. Mukundan, S. Madria, M. Linderman, A. N. Kang, L. Barolli, J. H. Park, Y.-S. Jeong, Y. Du, R. Zhang, M. Li, D. Chen, H. Zhao, H. Banirostam, A. Hedayati, S. K. Abd, S. a R. Al-Haddad, F. Hashim, and A. Abdullah, "Research on a security mechanism for cloud computing based on virtualization," 2012 Int. Conf. Comput. Sci. Electron. Eng., vol. 1, no. 4, pp. 19–24, 2013.

[10] D. Chen and H. Zhao, "Data Security and Privacy Protection Issues in Cloud Computing," The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012), vol. 1, pp. 647–651, 2012.

[11] A. Abuhussein, H. Bedi, and S. Shiva, "Evaluating Security and Privacy in Cloud Computing Services: A Stakeholder's Perspective," 2012 International Conference forInternet Technology And Secured Transactions, pp. 388–395, 2012.

[12] M. Almorsy, J. Grundy, and A. S. Ibrahim, "Collaboration-based cloud computing security management framework,"4th International. Conference Cloud ComputingIEEE, pp. 364–371, 2011

[13] A. Singh and H. Singh, "An improved LSB based image steganography technique for RGB images,"2015 IEEE International Conference onElectrical, Computer and Communication Technologies (ICECCT), pp. 1–4, 2015.

[14] H. Banirostam and a Hedayati, "A Trust Based Approach for Increasing Security in Cloud Computing Infrastructure," UKSim 15th IEEE Int. Conf. Compuer Model. Simul., pp. 717–721, 2013.