International Journal of Advanced Trends in Computer Applications
www.ijatca.com

Enhancing Privacy in Cloud Storage Using Efficient K-mean Clustering Technique

[1]Harmanpreet Kaur Sidhu, [2]Er. Mamoon Rashid
[1]Research Scholar
Department Of Computer Science Engineering
Chandigarh University, Gharaun, India
harman161991@gmail.com
[2]Assistant Professor
Department Of Computer Science Engineering
Chandigarh University, Gharaun, India.
mamoon873@gmail.com

Abstract: *Cloud computing is a model that enables convenient and on demand network access to a shared pool of configurable computing resources where millions of users share an infrastructure. Privacy and Security is significant obstacle that is preventing the extensive adoption of the public cloud in the Industry. Multi-tenancy where multiple tenants share cloud infrastructure poses an additional concern about the deliberate or accidental exposure of data. In this paper, the authors preserved privacy of data storage through k-mean clustering algorithm, IP addressing based detection technique, Longitude and latitude technique and cryptography technique for data uploading on cloud.*

Keywords: *Computing, Multi-tenancy, IP addressing, K-mean clustering, Cryptography*

I. Introduction

Cloud computing has begun to emerge as a hotspot in both industry and academia; It represents a new business model and computing paradigm, which enables on demand provisioning of computational and storage resources. Economic benefits consist of the main drive for cloud computing due to the fact that cloud computing offers an effective way to reduce capital expenditure and operational expenditure. The definition of cloud computing as per the literature in [1] is ”A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.”

The significance of cloud computing is arising and it is increasingly receiving a growing attention among academy, industry and scientific research. The cloud computing offers several benefits to small and intermediate enterprises as a utility computing because of less initial expenditure and even it opens the horizons for Information Technology [2]. It provides several services for data processing, storage, backup, facilitates productivity, fault tolerance, high scalability, high availability and communication etc [3]. But whenever a user shares information in cloud, the question of privacy and confidentiality comes in mind immediately.

Cloud computing is a network-based environment that focuses on sharing of computations or resources. Each user uses cloud services according to their subscription and according to the need of security requirement for their data. Cloud provider grants the resources and security according to the user’s need [4]. In multi-tenant environment, when the user moves information over the cloud they may lose control of it and data may accessed by the intruder. The main objective is to ensure that the data required by customer is not being accessed by or not being disclosed to any unauthorized person on the cloud [5].

Furthermore, we are dealing with huge amount of data and data mining is very useful to scrutinize the collected data .Moreover, it is not easy to analyze this data in raw form so small groups of similar data is formed with the help of clustering techniques when they are similar according to the some parameters [6].Clustering can be used in different ways according to the need of categorization. It is very frequently used method to save time for analyzing the large number of users in cloud and similar subscription users are grouped together in this paper. Different algorithms are K-mean, fuzzy c-means, mountain, subtractive etc [7].

This paper deals with the privacy problem which is due to the multi-tenant environment of cloud. The privacy is preserved by analyzing the logs of each user and if any unconditional activity or request is observed then that request is blocked if it fails to full fill the different security levels.

The rest of the paper is organized as follows: Section II describes the related work on privacy and security of data storage on cloud .In Section III, the problem is proposed by applying k-mean clustering technique and other security techniques for storing data on cloud.. Finally, Section IV discusses the results and conclusions drawn from the proposed work.

II. RELATED WORK ON PRIVACY OF DATA STORAGE:

In recent years, several methods have been proposed to preserve privacy and provide efficient analysis to the shared data.

Some interesting security issues are discussed in [8]. In this paper, the authors survey and analyze security, privacy and trust issues in cloud computing environments. An extensive review on cloud computing with the main focus on security gaps are presented in [9]. While in [10], Z. Dimitrios and L. Dimitrios attempt to evaluate cloud computing security by eliminating unique threats and introducing a trusted third party.

In [11], security risks of Multi Tenant architecture are addressed with two main risk factors. First risk is regarding to the virtualized infrastructure and other is related to the poorly implemented access management process and accidentally expose one user’s data to other users.

Mary et al. [12] proposed technique called Ant Colony Optimization (ACO) to improve k-means clustering algorithm.This paper has contributed to improve the quality of cluster after grouping. The proposed method consists of two phases.In the first phase, initial centroids are selected based on the statistical modes. In the second phase, the quality of the cluster is improved using the ant refinement algorithm.

G.Logeswari et. al [13], privacy of the shared PHR’s is preserved through data anonymization and encryption algorithm. In this efficient K-mean algorithm is purposed for clustering technique and it provides privacy to the through anonymization.

III. PROBLEM FORMULATION:

The proposed approach aims to enhancing the privacy on cloud storage using Efficient K-mean clustering algorithm, IP addressing based detection technique, Longitude and latitude technique and Cryptography technique for data uploading on cloud.

Figure 1 shows the components of this approach. The main component is Administrator. Administrator is responsible for the management of users, Data uploading and analyze the security on cloud through cryptography encryption algorithm.

Administrator creates users which are based on the subscription type whether it have limited access and unlimited access. Each user have their own subscription type and respective permissions according to subscription.

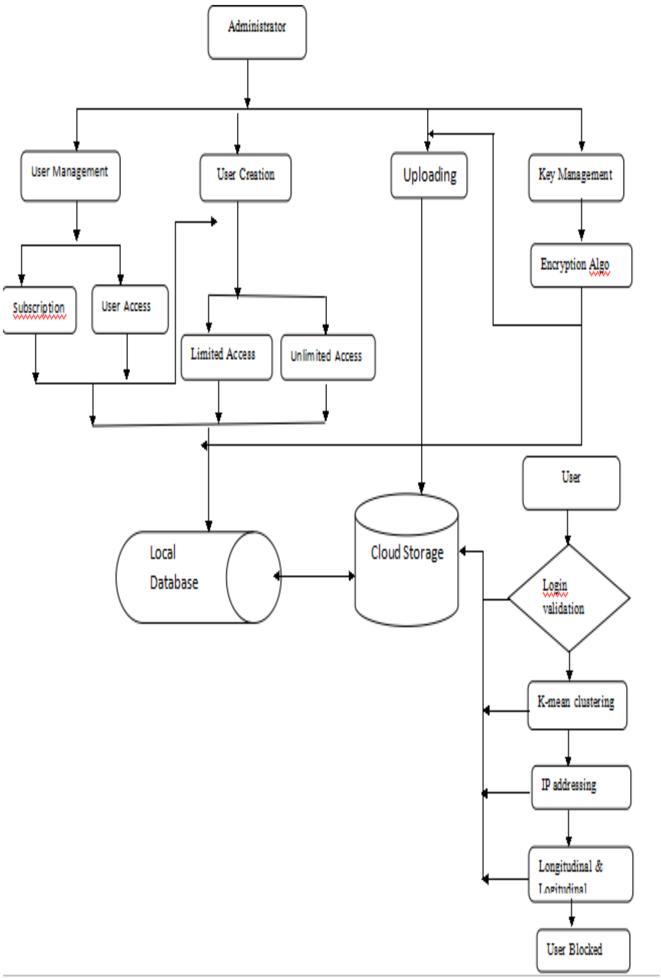

Fig. 1 Flow Diagram of Data Storage on Cloud with k-mean Clustering Technique

Second main task performed by administrator is uploading of data and give access to the data to users and assign permissions to the particular data so that users with different subscriptions can use them accordingly. Data is encrypted and uploaded on the cloud so that users can access them from anywhere, anytime if they are authenticated.

User authentication is the first step when user wants to access the data. If the user login for first time then it will directly access the data from cloud after validation. If it's not the first time and user want to download a file from the cloud storage then first K-mean clustering algorithm execute. After the execution of algorithm cluster of different users are formed according to their subscription type. Cluster of similar users are formed. After clustering the concept of privacy comes. If the user access the data within its permission then it will access the data from cloud otherwise it will move one step further for concerning the privacy of storage on cloud through IP addressing based detection technique. Logs are generated for each user in which IP address of the machine is also recorded through which user login .If IP address matches then user access the file otherwise moves to the next step which is longitudinal and latitudinal technique in which distance between the different IP address is calculated and if it is feasible to move from that place to another after the last login time then user is granted to access the file and if this will be failed then user is blocked and not able to access the cloud storage. When User clears all the privacy concerned steps then cryptographic algorithm is applied.

4. RESULTS AND CONCLUSION:

In this paper, first the authors proposed a new architecture for cloud data storage in which the private cloud should store only the organization's sensitive structure information and the public cloud should store the actual data in the cipher-text form by using newly designed encryption algorithm. The cryptographic algorithm will work on individual multimedia files for which every time the user will have to dynamically decrypt by using different keys. Moreover the authors have used K-Mean clustering technique for validating the user accesses based on spending subscriptions. The authors believe that the proposed system has the potential to be useful in commercial situations as it captures practical access policies based on roles in a flexible manner and provides secure data storage in the cloud enforcing these access policies.

References

[1] P.Mell and T. Grance, "The NIST Definition of Cloud Computing", 2011.

[2]Services in the Cloud Computing Era: A Survey

[3] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica,et al., "Above the clouds: A Berkeley view of cloud computing," University of California, Berkeley,Tech. Rep, 2009.

[4] Collaborative Network Security in Multi-Tenant Data Center for Cloud Computing

[5] Cloud Computing Architectures Based Multi-Tenant IDS

[6] A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis

[7]Khaled Hammouda and Fakhreddine Karray, "A Comparative Study of Data Clustering Techniques," in Tools of Intelligent Systems Design,2000.

[8] Dawei Suna, Guirau Chaugb, Lina Suna, Xingwei Wanga, "Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments", Procedia Engineering, Volume 15, Pages 2852-2856, 2011.

[9] Tanzim Khorshed, Shawkat Ali, Saleh A. Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing", Future Generation Computer Systems, Volume 28, Issue 6, Pages 833-851, June 2012.

[10] Dimitrios Zissis, Dimitrios Lekkas, "Addressing cloud computing security issues", Future Generation Systems, Volume 28, Issue 3, Pages 583-592, March 2012.

[11] "Securing Multi-tenaucy and Cloud Computing Security that ensures tenants do Not pose a risk to one another In terms of data loss, Misuse, or privacy Violation", Juniper Networks, Inc, Mar 2012.

[12]C.I. Mary, and S.V.K. Raja, "Refinement of clusters from k-means with ant colony optimization," Journal of Theoretical and applied Information Technology, 9(2):28-32, 2009.

[13]A Cost Effective Clustering based Anonymization Approach for Storing PHR's in Cloud.