Protection of the Privacy of the Cloud Data through Multiple Clouds

Amruta kammar[1] Vivekanandreddy[2] Navaneetha D[3] Sowmyashree B[4]

VTU, Belgaum

Abstract— The main goal of the “Protection of the privacy of the cloud data through multiple clouds” is to identify and understand the risks associated with the cloud data and propose a distributed architecture to eliminate the risks. In particular, in current cloud architecture a client entrusts a single cloud provider with his data. It gives the provider and outside attackers having unauthorized access to cloud, an opportunity of analyzing client data over a long period to extract sensitive information that causes privacy violation of clients. Hence to overcome this we propose a distributed architecture. The distributed approach to the cloud eliminates the disadvantage of storing all data of a client to the same provider. In this approach the data sent by the client is split into different number of chunks and is distributed among multiple cloud providers. Our system consists of two major components: Cloud data distributor and Cloud providers. The Cloud Data Distributor receives the data in the form of files from many clients, splits each of the files into much number of chunks and distributes these chunks among appropriate cloud providers. Another major component which is Cloud Providers, store these chunks and responds to the chunk requests by providing the chunks.

Keyword: - clouds; data mining; data privacy; security;

I. INTRODUCTION

Cloud computing has revolutionized the way computing and software services are delivered to the clients on demand. It provides users the ability to access to the computing resources and IT managed services with a previously unknown level of ease. Cloud computing provides end-users or small companies with facilities to use computational resources such as software, storage, and processing capacities which belongs to some other companies (cloud service providers). Cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

Though cloud computing is considered to be a powerful means of achieving high storage and computing services at a lower cost, it has not lived up to its reputation. Cloud has several security issues which involve assurance and confidentiality of data .A user who trusts the cloud provider for the security of the data may lose access to his data either temporarily or permanently because of an unlikely event such as a malware attack or network outage. Such kind of events can significantly harm the user’s data.

Confidentiality of user data is another big concern in the cloud. Cloud has been giving an opportunity to the providers to analyze the user data for a certain period of time. In addition, outside attackers who manage to get access to the cloud can also analyze data and violate user privacy. Cloud is not only a source of massive static data, but also a provider of high processing capacity at low cost. This makes cloud more vulnerable as attackers can use the raw processing power of cloud to analyze data. Various data analysis techniques are available now-a-days that successfully extract valuable information from a large volume of data. These analysis techniques are being used by cloud service providers.

Categorization allows to identify sensitive data and to take proper initiatives to maintain privacy of such data. Fragmentation and distribution of data among providers reduce the amount of data to a particular provider and thus minimize the risk associated with information leakage by any provider. This distribution is done according to the sensitivity of data and the reliability of cloud providers. The reliability of a cloud provider is defined in terms of its reputation. A cloud provider is given a particular data chunk only if the provider is reliable enough to store chunks of such sensitivity. Distribution restricts an attacker from having access to a sufficient number of chunks of data and thus prevents successful extraction of valuable information via mining. Even if an attacker manages to access required chunks, mining data from distributed sources remains a challenging job.

In this paper, we present an approach to prevent data mining based attacks on the cloud. Our system involves distributing user data among multiple cloud providers to make data mining a difficult job to the attackers. The key idea of our approach is to categorize user data, split data into chunks and provide these chunks to the appropriate cloud providers. In a nutshell our approach consists of categorization, fragmentation and distribution of data. The categorization of data is done according to mining sensitivity. Mining sensitivity in this context refers to the significance of information that can be leaked by mining.

II. LITERATURE SURVEY

In publication [1] the authors have presented a MapReduce based system known as Airavat. This system provides strong security and also guarantees privacy for distributed computations on sensitive data. Airavat is a system which provides end-to-end confidentiality, integrity and privacy guarantees along with a combination of access control and differential privacy. The security policies for the user’s sensitive data is controlled by the data providers by including mathematical bound on violation of privacy. Users can perform computations on their data without any security expertise. On behalf of users Airavat takes care of the computations by preventing the leakage of information beyond the data provider’s policy. The authors here have demonstrated the flexibility of Airavat on several case studies. The prototype which they have implemented is efficient with run times on Amazon’s cloud computing infrastructure within 32% of a Map Reduce system with no security.

Differential privacy is a new methodology ensures that the aggregate computation does not violate the privacy of individual inputs. This methodology adds some random noise to the output of computation. This noise has only a minor impact on computation’s accuracy. The authors here have presented a mathematically rigorous method for declassifying data in a mandatory access control system.

Program analysis techniques can be used to estimate how much information is leaked by a program. Privacy in MapReduce computations however is difficult if it is not possible to express it in a quantitative information flow problem. The flow bound cannot be set at 0 bits because the output depends on every single input. But even a 1-bit leakage may be sufficient to reveal, for example, whether a given person's record was present in the input dataset or not, violating privacy. Even though the differential privacy ensures the privacy violations of individual inputs, it cannot be specific to any given inputs.

In publication [2] the authors have discussed about the security and privacy issues at different levels viz. network, application and virtualization, in a cloud computing environment and proposed a security framework based on one-time pass key mechanism. The proposed security protocol provides security to both the service providers as well as the users in a highly conflicting cloud environment. The successful and secure extraction of useful information through data mining and clustering API depends on two major factors: Authenticity and Security.

To make the mechanism secure authors have proposed two types of security measures. First one is the use of "The One Time Password System" as pass key for authentication of API user. Second is the implementation of Cloud Service User Authentication Agent at the Server Side to authenticate the API user and client host details. The proposed approach integrates CLOUD API User IP authentication along with One Time Key based User Authentication by discarding malicious users from the domain reducing unauthorized access of API. It

also increases the security overhead using User's ID based on "One Time Pass Key" using Hashing principle and this framework fails to prevent malicious activity using any malicious code or parameter transfer procedure. Although the authors here have proposed the above mechanism to provide security and privacy to the user authentication, they have failed to identify the fraudulent activities which make the system to generate an alert about unauthorized fraudulent activity and also to overcome unauthorized data access and prevention in a heterogeneous cloud computing platform.

Therefore looking into the literature survey we can summarize that single cloud architecture is more vulnerable to security issues because data remains under the single cloud. Hence to overcome this we present an multi-cloud architecture which eliminates all these issues and provides an extra protection to the users data with the help of encryption.

III. Data Mining On Cloud

Data mining is one of the fastest growing fields in computer industry that deals with discovering patterns from large data sets. It is considered to be a part of the knowledge discovery process and is used to extract the information which is understandable by the humans. Mining is usually applied on large amount of data and to create qualitative models from these huge data we require related algorithms which are helpful in extracting the useful information from this raw data.

The relationship between data mining and cloud is worth to discuss. With the help of a data mining cloud providers provides a better service to their clients. The issues like privacy and individuality will be violated if the clients are unaware of the information being collected by the unauthorized users. This can be a serious privacy issue if both the cloud providers as well as outside attackers misuse the information. The outside attackers are the one who doesn't have authorized access to the cloud. These attackers can make use of the cheap and raw computing power which is provided by the cloud computing to extract the useful information from the huge amount of data.

A. *The importance of client privacy*

Client privacy is a serious issue as not all clients have the same demands regarding privacy of their data. Some clients are satisfied with the policy provided by the particular cloud while other clients are more concerned much about their privacy. The proposed system is designed especially for the client who falls under the second category and for whom the privacy is of great concern. These clients may not afford the luxury of maintaining private storage while they are interested in spending a little more money on maintaining their privacy on the cloud.

B. *Data Mining: A Potential Threat to Privacy*

The successful extraction of useful information via data mining depends on two main factors: proper amount of data and suitable mining algorithms. Various mining algorithms are used for numerous purposes. Some mining algorithms are good enough to extract information up to the limit that violates client privacy.

IV. A Distributed Approach Eliminating Cloud Mines

In this section, we first discuss the existing system threats to the single provider cloud architecture. Then give an overview of the distributed approach which helps to eliminate the privacy risks identified in the existing systems.

A. *Existing System Threats*

The current cloud storage system is a vulnerable one because data remain under a single cloud provider. This can lead to data loss in case of events like network outage, the cloud provider going out of business, malware attack etc. This eases the job of the attackers. As long as the entire data belonging to one client remain under a single cloud provider, both inside and outside attackers gets the benefit of using data mining to a great extent. Inside attacker is the malicious employees at a cloud provider and outside attacker is the hacker. Thus single provider architecture is the biggest security threat concerning data mining on cloud.

B. *Our Proposed System*

The distributed approach to the cloud eliminates the disadvantage of storing all data of a client to the same provider. The data (file) is split into different number of chunks and is distributed among multiple cloud providers. Our system consists of two major components: Cloud data distributor and Cloud providers. The Cloud Data Distributor receives the data in the form of files from many clients, splits each of the files into many numbers of chunks and distribute these chunks among appropriate cloud providers. Another major component which is Cloud Providers, store these chunks and responds to the chunk requests by providing the chunks. The clients can retrieve the data by downloading the uploaded file. It also gives the option of immediately viewing the file or saving the file for later use. The distributed architecture for cloud redefines the partitioning of data in terms of preserving privacy.

V. System Overview

In this section we discuss our proposed system architecture that prevents the privacy attacks on the cloud. Our system consists of two major components: Cloud Data Distributor and Cloud Providers. The Cloud Data Distributor receives data in the form of files from clients, splits each file into chunks and distributes these chunks among cloud providers. Cloud Providers store chunks and responds to chunk requests by providing the chunks.

A. Cloud Data Distributor

Cloud Data Distributor is the entity that receives data (files) from clients, performs fragmentation of data (splits files into chunks) and distributes these fragments (chunks) among Cloud Providers. It also participates in data retrieving procedure by receiving chunk requests from clients and forwarding them to Cloud Providers. Clients do not interact with Cloud Providers directly rather via Cloud Data Distributor. This entity deals with Cloud Providers as an agent of clients.

To upload the data, the clients have to login through their user-id and password. After the successful login users are provided the facility of uploading the files, Downloading the files and removing the unwanted file from the cloud. While uploading the data to the cloud data distributor each file is given a different sensitivity level such as high and low. According to the sensitivity level selected the split file is sent to the appropriate cloud providers

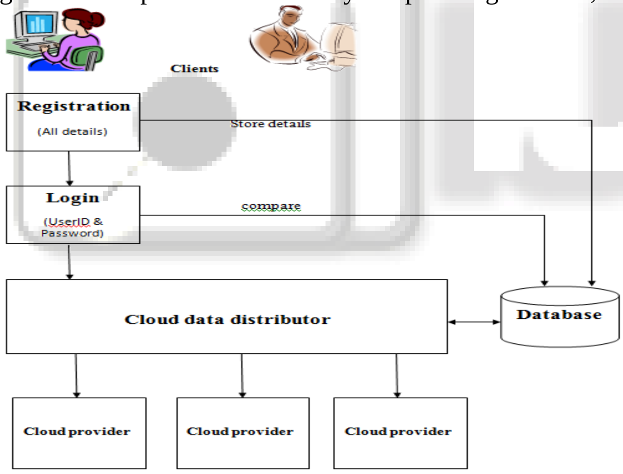


Fig. 1: System Architecure

B. Cloud Providers

The second entity refers to the cloud storage providers. The main tasks of Cloud Providers are: storing chunks of data, responding to a query by providing the desired data, and removing chunks when asked. Providers receive chunks from the distributor and store them. Each provider is considered as a separate disk storing clients' data. The cloud provider responds to the query of the distributor by providing data.

VI. Implementation

Modular Specification

1. Registration
2. Login to cloud
3. Distribution of data
4. Retrieval of data
5. Removal of data

To implement our proposed system, we need to implement the following functionalities.

A. Registration

This module involves in the registration of the new users. The users who want to get access to cloud have to first register themselves by providing their details such as user-id, password, name of the user, address of the user and mobile number of the user. These details will be stored in the database for later access. Only the given password is stored in the encrypted format. Further the users can unregister by providing their user-id. The corresponding details belonging to that particular user-id is deleted permanently from database. The user no longer has access to cloud.

B. Login to cloud

This module involves in the authentication of already registered users. The users are required to provide their user-id and password. The password given by the user is encrypted and compared against the encrypted password in the database. If both the password matches then the user is an authorized user and he is redirected to carry out further tasks such as uploading, retrieving and removing of files.

C. Data distribution

The data provided by the clients is split into number of chunks. The chunks are formed depending on the size of the file. Further these chunks are distributed among multiple providers to achieve the privacy of the data.

D. Data Retrieval

This module retrieves the already uploaded file from the particular cloud. This can be done by providing the name of the file and the corresponding password given to the file.

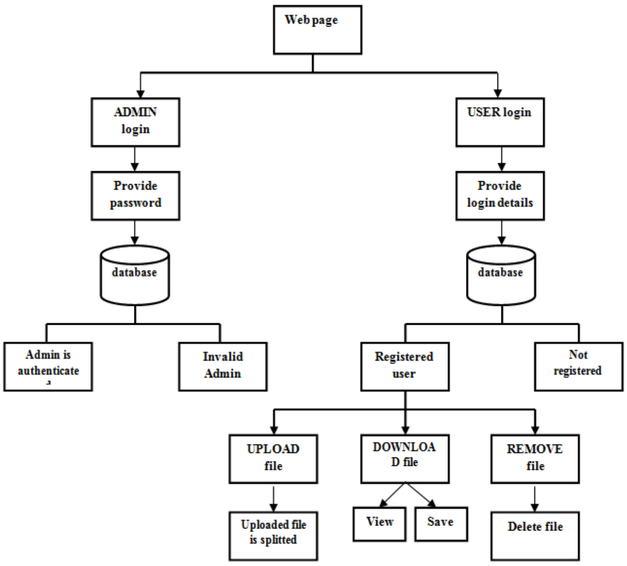


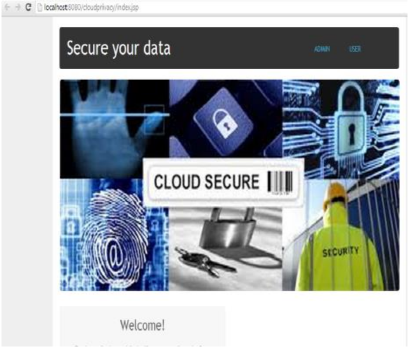
Fig. 2: DFD of the application

E. Data Removal

This module accepts a file name and the corresponding user-id of the client. After accepting the details files are listed for that particular user. Further the user can delete unwanted files from his list.

VII. EXPERIMENTAL RESULTS

We have implemented our proposed system using java modules. The tool used to implement our proposed system is Net beans 6.0.1. We have implemented cloud data distributor and cloud providers.

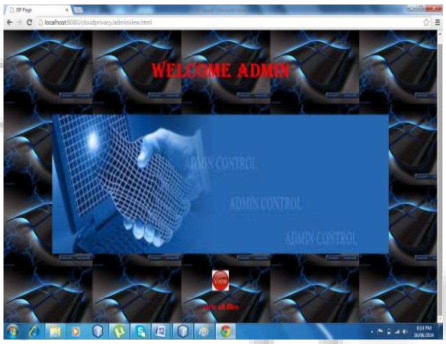

localhost:8080/cloudprivacy/index.jsp

Secure your data

Welcome!

Fig. 1: Display of webpage of the application

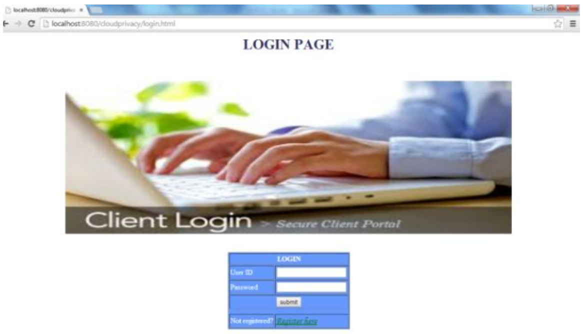
The admin is provided with the facility of viewing files of users.


localhost:8080/cloudprivacy/adminview.html

WELCOME ADMIN

Fig. 2: Admin control page

The users are provided with the facility of registration, login, UPLOAD/DOWNLOAD /REMOVE files.

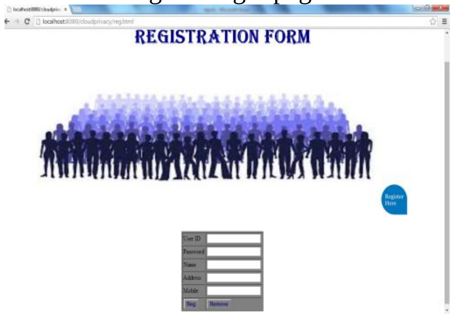

localhost:8080/cloudprivacy/login.html

LOGIN PAGE

Client Login > Secure Client Portal

LOGIN	
User ID	
Password	
	submit
Not registered?	Register here

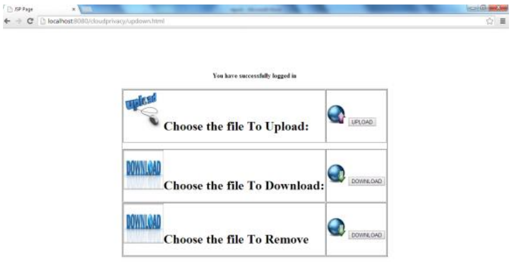
Fig. 3: Login page


localhost:8080/cloudprivacy/reg.html

REGISTRATION FORM

User ID	
Password	
Name	
Address	
Mobile	
Reg	Remove

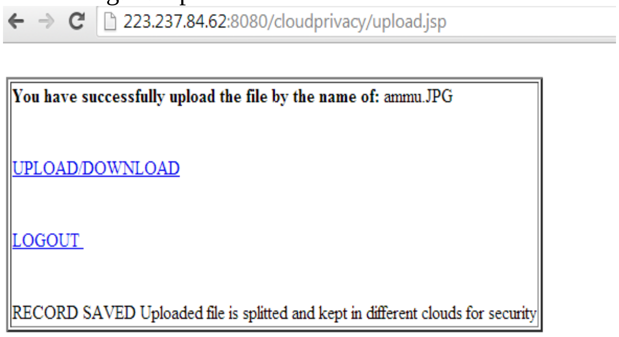
Fig. 4: Registration of users


localhost:8080/cloudprivacy/updown.html

You have successfully logged in

Choose the file To Upload:	UPLOAD
Choose the file To Download:	DOWNLOAD
Choose the file To Remove	DOWNLOAD

Fig. 5: Upload/Download/Remove of files


223.237.84.62:8080/cloudprivacy/upload.jsp

You have successfully upload the file by the name of: ammu.JPG

UPLOAD/DOWNLOAD

LOGOUT

RECORD SAVED Uploaded file is splitted and kept in different clouds for security

Fig. 6: Results of file upload

VIII. CONCLUSION AND FUTURE WORK

Ensuring security of cloud data still a challenging problem. Cloud service providers as well as other third parties use different data mining techniques to acquire valuable information from user data hosted on the cloud. In this project, we have discussed the impact of data mining on cloud and have proposed a distributed structure to eliminate mining based privacy threat on cloud data. Our approach combining categorization, fragmentation and distribution, prevents data mining by maintaining privacy levels, splitting data into chunks and storing these chunks of data to appropriate cloud providers.

Although the proposed system provides an effective way to protect privacy from mining based attacks, it introduces performance overhead when client needs to access all data frequently, e.g. client needs to perform a global data analysis on all data. The analysis may have to access data from multiple locations, with a degraded performance. In future, we look forward to improve our system by reducing such overhead.

REFERENCES

[1] Indrajit Roy, Srinath T.V. Setty, Ann Kilzer, Vitaly Shmatikov, Emmett Witchel, "Airavat: Security and Privacy for MapReduce".

[2] Rohit Bhadauria*, Raj deep Borgohain, Abirlal Biswas, Sugata Sanyal, "Secure Authentication of Cloud Data Mining API".

[3] Tingting Hu , Haishan Chen , Xiaodan Zhu, Lu Huang, "A Survey of Mass Data Mining Based on Cloudcomputing".

[4] Himel Dev, Tanmoy Sen, Madhusudan Basak and Mohammed Eunus Ali , "An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks", Published in IEEE computer society,2013.

[5] Ruxandra-Ştefania PETRE, "Data mining in Cloud Computing", Published in Database Systems Journal vol. III, no. 3/2012.