

# **ATTRIBUTE-BASED STORAGE SUPPORTING SECURE DEDULICATION OF ENCRYPTED DATA IN CLOUD**

**A PROJECT REPORT**

*Submitted to*

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**

*In partial fulfilment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

*by*

**Reddem Jashwanth Kumar Reddy**

**191611059**

*Supervisor*

**Dr.J.Rene Beulah**



**SAVEETHA SCHOOL OF ENGINEERING**

**SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL**

**SCIENCES, CHENNAI – 602 105**

**June 2020**

## **BONAFIDE CERTIFICATE**

Certified that this project report "**ATTRIBUTE-BASED STORAGE SUPPORTING SECURE DEDUPLICATION OF ENCRYPTED DATA IN CLOUD**" is the bonafide work of "**REDDEM JASHWANTH KUMAR REDDY (Reg. No. 191611059)**" who carried out the project work under my supervision.

**SIGNATURE**

**Dr. SP. Chokkalingam**

**HEAD OF THE DEPARTMENT**

Professor, Dept. of CSE

Saveetha School of Engineering

SIMATS, Chennai - 602 105

**SIGNATURE**

**Dr.J.Rene Beulah**

**PROJECT SUPERVISOR**

Designation, Dept. of CSE

Saveetha School of Engineering

SIMATS, Chennai – 602105

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

## **DECLARATION BY THE CANDIDATE**

I declare that the report entitled **“Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud”** submitted by me for the degree of Bachelor of Engineering is the record of the project work carried out by me under the guidance of **“Dr.J.Rene Beulah”** and this work has not formed the basis for the award of any degree, diploma, associateship, fellowship, titled in this or any University or other similar institution of higher learning.

**SIGNATURE**

**R.Jashwanth kumar Reddy**  
**(Reg. No. 191611059)**

## ABSTRACT

Attribute-based encryption (ABE) has been widely used in cloud computing where a data provider outsources his/her encrypted data to a cloud service provider, and can share the data with users possessing specific credentials (or attributes). However, the standard ABE system does not support secure deduplication, which is crucial for eliminating duplicate copies of identical data in order to save storage space and network bandwidth. In this paper, we present an attribute-based storage system with secure deduplication in a hybrid cloud setting, where a private cloud is responsible for duplicate detection and a public cloud manages the storage. Compared with the prior data deduplication systems, our system has two advantages. Firstly, it can be used to confidentially share data with users by specifying access policies rather than sharing decryption keys. Secondly, it achieves the standard notion of semantic security for data confidentiality while existing systems only achieve it by defining a weaker security notion. In addition, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under other access policies without revealing the underlying plaintext.

## ACKNOWLEDGEMENT

This project work would not have been possible without the contribution of many people. It gives me immense pleasure to express my profound gratitude to our honorable Chancellor, **Dr. N. M. Veeraiyan**, Saveetha Institute of Medical and Technical Sciences, for his blessings and for being a source of inspiration. I sincerely thank our Vice Chancellor, **Dr. Rakesh Kumar Sharma**, for his visionary thoughts and support. I am indebted to extend my gratitude to our Director Madam, **Mrs. Ramya Deepak**, Saveetha School of Engineering, for facilitating us all the facilities and extended support to gain valuable education and learning experience.

I register my special thanks to **Dr. D. Dhanasekaran**, Principal, Saveetha School of Engineering and **Dr. SP.Chokkalingam**, HoD, Department of Computer Science and Engineering, for the support given to me in the successful conduct of this project. I wish to express my sincere gratitude to my supervisor, **Dr.J.Rene Beulah** for her inspiring guidance, personal involvement and constant encouragement during the entire course of this work.

I am grateful to Project Coordinators, Review Panel External, Internal Members and the entire faculty of the Department of Computer Science and Engineering, for their constructive criticisms and valuable suggestions which have been a rich source to improve the quality of this work.

**R.Jashwanth kumar Reddy**

<b>CHAPTER</b>	<b>TITLE</b>	<b>PAGE</b>
	<b>ABSTRACT</b>	<b>IV</b>
	<b>LIST OF FIGURES</b>	<b>VII</b>
<b>1</b>	<b>INTRODUCTION</b>	
	1.1. INTRODUCTION OF DOMAIN	1
	1.2. OBJECTIVES	8
	1.3. SCOPE OF PROJECT	9
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>11</b>
<b>3</b>	<b>PROBLEM STATEMENT AND METHODOLOGY</b>	
	3.1 PROBLEM DEFINITION	23
	3.2.1 EXISTING SYSTEM	23
	3.2.2 PROPOSED SYSTEM	24
	3.3 TECHNIQUES	24
	3.4 SYSTEM ARCHITECTURE	26
	3.5 UML DIAGRAMS	27
	3.5.1 USE CASE DIAGRAM	27
	3.5.2 CLASS DIAGRAM	28
	3.5.3 ACTIVITY DIAGRAM	28
	3.5.4 DEPLOYMENT DIAGRAM	29

	3.5.5 COMPONENT DIAGRAM	29
<b>4</b>	<b>SYSTEM IMPLEMENTATION</b>	
	4.1. MODULE DESCRIPTION	30
	4.2 ALGORITHM OF PROPOSED WORK	31
	4.3 FLOWCHART	32
	4.4 DESCRIPTION OF DATASET	33
	4.5 COMPARATIVE STUDY OF EXISTING AND PROPOSED SYSTEM	34
<b>5</b>	<b>RESULTS AND DISCUSSION</b>	
	5.1 OUTPUT SCREENSHOTS WITH EXPLANATION	35
	5.2 ANALYSIS OF OUTPUT	39
<b>6</b>	<b>CONCLUSION AND FUTURE SCOPE</b>	
	6.1. CONCLUSION	40
	6.2 FUTURE SCOPE	40
	<b>REFERENCES</b>	42
	<b>ANNEXURE I - SAMPLE CODE</b>	44
	<b>ANNEXURE II – PUBLICATION COPY</b>	
	<b>APPENDIX III – CD</b>	

## LIST OF FIGURES:

FIGURE NO.	TITLE	PAGE NO.
3.3	System Architecture	27
3.4	UML Diagrams	28
3.4.1	Use case Diagram	28
3.4.2	Class Diagram	29
4.3	Flow diagram	34
5.1.1	Output Screenshot 1	36
5.1.2	Output Screenshot 2	37
5.1.3	Output Screenshot 3	37
5.1.4	Output Screenshot 4	38
5.1.5	Output Screenshot 5	38
5.1.6	Output Screenshot 6	39



## **List of Abbreviations**

CSP – Cloud Service Provider

ABE – Attribute based Encryption

ECC - Elliptic Curve Cryptography

IaaS - Infrastructure as a Service

PaaS - Platform as a Service

SaaS - Software as a Service

## **1. INTRODUCTION**

### **1.1 Introduction of Project**

Cloud Computing offers a new way technology by arranging resources of various types like computing and storage, based on the demands cloud provide information to the users or clients. It has desirable features like elasticity, fault-tolerance, security, encryption, pay as you use model. Promising platform for clients to store and manage a data in remote server where user can access from anywhere. The better approach for data innovation benefit offered by distributed computing is modifying different assets and the information is given to clients on their requests. This is turned into a promising administration stage due to a few properties. For example, adaptation to non-critical failure, pay per utilize versatility, and flexibility are the alluring properties of Cloud Computing. The clients of cloud transfer secret or individual information to the datacentre of CSP (Cloud Service Provider such as Amazon, Google and so on.) and enable it to keep up these information. Because of a few assaults and interruption towards touchy information at CSP are not avoidable. Cloud clients can't completely believe the CSP. The security issue turns out to be more genuine because of alternate investigation innovations and the quick improvement of Data Mining. Some of the time the deduplicated information in encoded frame to CSP might be transferred by same or distinctive cloud clients. Putting away similar information in scrambled frame or ordinary information or Data deduplication squanders assets of system, entangles the administration, part of vitality devours. For the information holders it is hard to keep up the deduplication because of many reasons. For example,

- 1) Storage deferral is brought about Data holders may not be in online dependably or accessible for such administration,

- 2) Deduplication turn out to be excessively confused as far as Computation and correspondence to include information proprietors into deduplication prepare.

- 3) the way toward finding the deduplication may barge in the security of information holders. Hence Cloud benefit furnishes can't coordinate with information holders on information stockpiling deduplication as a rule. High cost saving is achieved and proved by Deduplication. Reducing upto 65% in file systems and 90-95 % storage needs backup applications. Existing

Systems are not able to deduplicate the encrypted data and cannot ensure security privacy authentication, reliability. When data holders are not online it's hard to manage the deduplication due to many reasons, and it causes storage delay. This paper works on encryption algorithms, to find which performs better. Using 4 algorithm such as ECC(Elliptic Curve Cryptography), DES(Data Encryption Standard), AES(Advanced Encryption Standard) and RSA. Data ownership challenges, digital signature, to manage data which is encrypted use PRE. Our goal is to solve data duplication problem and to save storage space in other way saving money. Already user saved the document in the cloud and when the other user try to save the same content with the different name, it should tell the second user that the content is already existing. In this work try to avoid Duplication to save Storage Space as the user is are paying for cloud its necessary to think about the storing space, user should not store same data more than one time, if user do that its waste of storage space in other words users are wasting our own money. Encrypted data introduce new challenges for cloud data de duplication and Traditional de duplication schemes cannot work on encrypted data. The deduplicated data in encrypted form to CSP may be uploaded by same or different cloud users. Storing the same data in encrypted form or normal data cause Data deduplication wastes resources of network, complicates the management, lot of energy consumes. For the data holders it is difficult to maintain the deduplication due to many reasons. Objective of this work is: To design and implement a solution to deduplicate the— encrypted big data in cloud. To increase the efficiency, effectiveness and— applicability. To save the storage space in cloud and protect the— privacy of data holders or cloud users. The solution can flexibly supports sharing the data even— when the data owner is not in online.

## **Introduction of Domain**

**Cloud computing** is the delivery of computing and storage capacity as a service to a heterogeneous community of end-recipients. The name comes from the use of cloud-shaped symbols an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts services with a user's data, software and computation over a network.

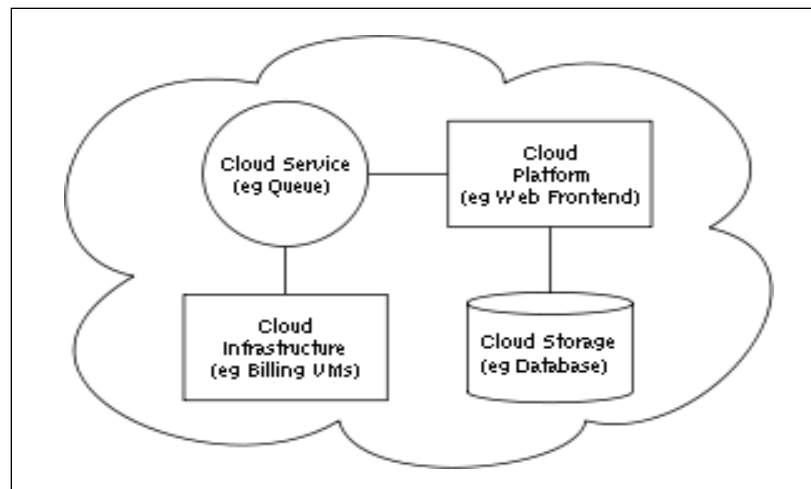
There are three types of cloud computing:

- Infrastructure as a Service (IaaS),
- Platform as a Service (PaaS), and
- Software as a Service (SaaS).

Using Infrastructure as a Service, users rent use of servers (as many as needed during the rental period) provided by one or more cloud providers. Using Platform as a Service, users rent use of servers and the system software to use in them. Using Software as a Service, users also rent application software and databases. The cloud providers manage the infrastructure and platforms on which the applications run.

## CLOUD COMPUTING ARCHITECTURE

Cloud architecture, the systems architecture of the software systems involved in the delivery of cloud computing, typically involves multiple cloud components communicating with each other over a loose coupling mechanism such as a messaging queue.



**Figure 1: Cloud computing sample architecture**

Elastic provision implies intelligence in the use of tight or loose coupling as applied to mechanisms such as these and others

## **Cloud computing types**

- Public cloud
- Community cloud
- Hybrid cloud
- Private cloud

## **Characteristics**

- Cost
- Reliability
- Flexibility
- Scalability
- Performance
- Security
- Maintenance
- Virtualization

Cloud computing is defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are harnessed to solve problems too intensive for any stand-alone machine. In cloud computing, the word cloud (also phrased as "the cloud") is used as a metaphor for "the Internet," so the phrase cloud computing means "a type of Internet-based computing," where different services — such as servers, storage and applications — are delivered to an organization's computers and devices through the Internet. Cloud Computing is a technology that uses the internet and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with internet access. This technology allows for much more efficient computing by centralizing data storage, processing and bandwidth. A simple example of cloud computing is Yahoo email, Gmail, or Hotmail etc. All you need is just an internet connection and you can start sending emails. The server and email management software is all on the cloud (internet) and is totally managed by the cloud service provider Yahoo , Google

etc. Cloud computing is broken down into three segments: "application" "storage" and "connectivity." Each segment serves a different purpose and offers different products for businesses and individuals around the world. In June 2011, a study conducted by V1 found that 91% of senior IT professionals actually don't know what cloud computing is and two-thirds of senior finance professionals are clear by the concept, highlighting the young nature of the technology. In Sept 2011, an Aberdeen Group study found that disciplined companies achieved on average an 68% increase in their IT expense because cloud computing and only a 10% reduction in data center power costs.

### **How Cloud Computing Works**

The goal of cloud computing is to apply traditional supercomputing, or high-performance computing power, normally used by military and research facilities, to perform tens of trillions of computations per second, in consumer-oriented applications such as financial portfolios, to deliver personalized information, to provide data storage or to power large, immersive online computer games. The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing chores across them. This shared IT infrastructure contains large pools of systems that are linked together. Often, virtualization techniques are used to maximize the power of cloud computing.

#### *Cloud Computing in the Data Center and for Small Business*

Cloud computing has started to obtain mass appeal in corporate data centers as it enables the data center to operate like the Internet through the process of enabling computing resources to be accessed and shared as virtual resources in a secure and scalable manner. For a small and medium size business (SMB), the benefits of cloud computing is currently driving adoption. In the SMB sector there is often a lack of time and financial resources to purchase, deploy and maintain an infrastructure (e.g. the software, server and storage). In cloud computing, small businesses can access these resources and expand or shrink services as business needs change. The common pay-as-you-go subscription model is designed to let SMBs easily add or remove services and you typically will only pay for what you do use.

## **1.2 Objective of the Problem**

There are several security issues that threaten the cloud environment by the above issues at the time of data retrieval and de-duplication. To protect the privacy of data and resist unwanted accesses in the cloud storage data. Because the storage data can be a sensitive data, which is categorized in a different format like health data, personal photos, secure tax contents etc., so, these contents should be encrypted by data owners before outsourcing to the cloud storage. While considering the content retrieval and data search, the traditional service is based on plain-text keyword search methods. The irrelevant process of downloading all the data and decrypting locally is clearly impractical and not cost effective. Due to a large amount of bandwidth cost in cloud scale systems, it creates a big trouble. The uploaded contents are sometimes treated as important information for the other category. Therefore, exploring privacy preserving and effective search service over encrypted cloud data is of great importance in the cloud storage service. Considering the potentially huge number of on-demand data users and a lot of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability.

## **1.3 Scope of the project**

Cloud computing confers strong economic advantages, but many clients are reluctant to implicitly trust a third-party cloud provider. To address these security concerns, data may be transmitted and stored in encrypted form. Major challenges exist concerning the aspects of the generation, distribution, and usage of encryption keys in cloud systems. To prevent unauthorized data usage, fine-grained access control is necessary in multi-user system. However, authorized user may intentionally leak the secret key for financial benefit. Thus, tracing and revoking the malicious user who abuses secret key needs to be solved imminently. In the recent trend, every data and contents are stored in the cloud using cloud storage services. With the huge amount of data from every client may affect the cloud storage. In specific, the redundant content may perform more worst in the storage part. The de-duplication method is generally used to reduce the storage cost and resource requirements of data services in the cloud by eliminating redundant data and storing only a single copy of them. De-duplication is most effective when multiple

users outsource the same data to the cloud storage services, but it creates several issues relating to search and security. Data mining is an effective way to solve such problems in the cloud service.



## **2. LITERATURE REVIEW**

### **2.1 Improving the Performance of System in Cloud by Using Selective Deduplication.**

**Nishant N. Pachpor ; Prakash S. Prasad IEEE 2018.**

Cloud Computing is very popular today because of large amount of data storage and fast access of data over the network. But in today' s scenario we find the some issue to access and store data in cloud likewise data theft, data loss, privacy issue, infected application, data location, security on vendor level, security at user level and data duplication. As we find of recent study 7 Zeta Byte (ZB) data available in different storage location after 5 years it will increases the 5 times more data storage. For the better performance of system we use the different data deduplication method liked selective performance oriented data deduplication. In this paper we propose to remove data redundancy from available offline or online data storage as well as we provide security of data which helps to improve the performance of system. After deleting the data from file automatically size of file reduces and which helps to reduce the traffic on the network.

### **2.2 A Big Data Deduplication Using HECC Based Encryption with Modified Hash Value in Cloud Ankit Shrivastava ; Abhigyan Tiwary IEEE 2018.**

Big data deduplication is one of the most challenging task in the cloud world. There are two major issue generated in the cyber world first is the data preservation on cloud and second one is big duplication. In this research proposed a new model to solve both problems. In this paper proposed modified hash value concept, with the help of this avoid big data problem and for secure data protection use HECC algorithm for data encryption and decryption. SHA2 algorithm consume less time as compare to SHA-1 for hash value generation and HECC shows better encryption as compare to other methods. In this research also analysed the different methods such AES, DSA and ECC for data encryption on the basic of time complexity. The proposed system shows better result as compare to other previous data duplication methods for the basis of time and security.

### **2.3 Design of Global Data Deduplication for a Scale-Out Distributed Storage System** **Myoungwon Oh ; Sejin Park ; Jungyeon Yoon ; Sangjae Kim ; Kang-won Lee ; Sage Weil ; Heon Y. Yeom ; Myoungsoo Jung IEEE 2018.**

Scale-out distributed storage systems can uphold balanced data growth in terms of capacity and performance on an on-demand basis. However, it is a challenge to store and manage large sets of contents being generated by the explosion of data. One of the promising solutions to mitigate big data issues is data deduplication, which removes redundant data across many nodes of the storage system. Nevertheless, it is non-trivial to apply a conventional deduplication design to the scale-out storage due to the following root causes. First, chunk-lookup for deduplication is not as scalable and extendable as the underlying storage system supports. Second, managing the metadata associated to deduplication requires a huge amount of design and implementation modifications of the existing distributed storage system. Lastly, the data processing and additional I/O traffic imposed by deduplication can significantly degrade performance of the scale-out storage. To address these challenges, we propose a new deduplication method, which is highly scalable and compatible with the existing scale-out storage. Specifically, our deduplication method employs a double hashing algorithm that leverages hashes used by the underlying scale-out storage, which addresses the limits of current fingerprint hashing. In addition, our design integrates the meta-information of file system and deduplication into a single object, and it controls the deduplication ratio at online by being aware of system demands based on post-processing. We implemented the proposed deduplication method on an open source scale-out storage. The experimental results show that our design can save more than 90% of the total amount of storage space, under the execution of diverse standard storage workloads, while offering the same or similar performance, compared to the conventional scale-out storage.

### **2.4 A Comparative Study of Data Deduplication Strategies** **Nipun Chhabra ; Manju Bala IEEE 2018.**

Data Deduplication is a new way of compressing data which is helpful in efficient use of storage space and proves to be a better technique to handle duplicate data. Deduplication allows unique data copy to be uploaded to the storage originally and successive copies are provided

with connecting pointer to the authentic stored copy. In cloud computing, data privacy is a significant matter to be considered. This paper surveys the prevalent deduplication strategies and evaluates their methodologies. The survey thus conducted is organized in tabular form to bring out the crisp of progression in the deduplication and its types. We have observed during this survey that in deduplication, confidentiality of data is being compromised at many levels so, it paves a path for a more secure approach in data deduplication in cloud computing.

**2.5 RARE: Defeating side channels based on data-deduplication in cloud storage Zahra Pooranian ; Kang-Cheng Chen ; Chia-Mu Yu ; Mauro Conti IEEE 2018.**

Client-side data deduplication enables cloud storage services (e.g., Dropbox) to achieve both storage and bandwidth savings, resulting in reduced operating cost and high level of user satisfaction. However, the deduplication checks (i.e., the corresponding essential message exchange) create a side channel, exposing the privacy of file existence status to the attacker. In particular, the binary response from the deduplication check reveals the information about the existence of a copy of the file in the cloud storage. This behavior can be exploited to launch further attacks such as learning the sensitive file content and establishing a covert channel. While current solutions provide only weaker privacy or rely on unreasonable assumptions, we propose RANdom REsponse (RARE) approach to achieve stronger privacy. The idea behind our proposed RARE solution is that the uploading user sends the deduplication request for two chunks at once. The cloud receiving the deduplication request returns the randomized deduplication response with the careful design so as to preserve the deduplication gain and at the same time minimize the privacy leakage. Our analytical results confirm privacy guarantee and results show that both deduplication benefit and privacy of RARE can be preserved.

**2.6 Privacy-Preserving and Updatable Block-Level Data Deduplication in Cloud Storage Services Hyungjune Shin ; Dongyoung Koo ; Youngjoo Shin ; Junbeom Hur IEEE 2018.**

To achieve high storage saving, data deduplication techniques are widely used in many practical cloud storage services, which removes redundant data and keeps only a single copy of them. However, secure data deduplication over encrypted data is challenging since encryption may result in different ciphertexts even when the original messages are the same. Thus, message-

locked encryption (MLE) is proposed to solve this issue and demonstrates that it is secure under the unpredictable message set. Since block-level deduplication can achieve more fine-grained storage saving, several block-level deduplication schemes that support updatability are also vividly proposed. However, the previous updatable block-level MLE schemes are vulnerable against brute-force attack when the message set is predictable. Since the size of a block is typically much less than an arbitrary size of a file, the predictability problem is a very important pragmatic concern which should be addressed in the block-level deduplication literature. In this paper, thus, we propose a novel secure block-level deduplication scheme that guarantees efficient data update and brute-force attack resilience even when messages are predictable with the rigorous security proof. Also, our performance evaluation shows that additional time and bandwidth usage can be minimized as the size of a block increases.

## **2.7 GDedup: Distributed File System Level Deduplication for Genomic Big Data Paul Bartus ; Emmanuel Arzuaga IEEE 2018.**

During the last years, the cost of sequencing has dropped, and the amount of generated genomic sequence data has skyrocketed. As a consequence, genomic sequence data have become more expensive to store than to generate. The storage needs for genomic sequence data are also following this trend. In order to solve these new storage needs, different compression algorithms have been used. Nevertheless, typical compression ratios for genomic data range between 3 and 10. In this paper, we propose the use of GDedup, a deduplication storage system for genomics data, in order to improve data storage capacity and efficiency in distributed file systems without compromising I/O performance. GDedup can be developed by modifying existing storage system environments such as the Hadoop Distributed File System. By taking advantage of deduplication technology, we can better manage the underlying redundancy in genomic sequence data and reduce the space needed to store these files in the file systems, thus allowing for more capacity per volume. We present a study on the relation between the amount of different types of mutations in genomic data such as point mutations, substitutions, inversions, and the effect of such in the deduplication ratio for a data set of vertebrate genomes in FASTA format. The experimental results show that the deduplication ratio values are superior to the actual compression ratio values for both (file read-decompress or write-compress) I/O patterns,

highlighting the potential for this technology to be effectively adapted to improve storage management of genomics data.

## **2.8 PTS-Dep:A High-Performance Two-Party Secure Deduplication for Cloud Storage.**

**Wenlong Tian ; Ruixuan Li ; Weijun Xiao ; Zhivong Xu IEEE 2018.**

In cloud storage, the message-locked encryption method is widely used in security deduplication. However, Brute force attack becomes a serious issue. Current research addresses the brute force attack problem in secure deduplication using a third-party model. Even though there is a trusted third party in real life, it is hard to be applied to traditional two-party based deduplication system which only includes the client and the storage provider. It is obvious that industries prefer to take the simpler and more practical secure architecture under the same level of security. However, the existing two-party secure deduplication approaches either have inferior performance or security holes. To make the two-party secure deduplication comparable in performance with unprotected baseline and keep the same level security with the existing two-party secure deduplication, we propose a high-performance two-party secure deduplication, PTS-Dep. By leveraging secure duplicate data detection scheme and secure duplicate data's key sharing scheme, PTS-Dep can perform data deduplication with the security guarantee. Our approach improves average deduplication performance up to 92\% for Fslhome workloads compared to previous secure deduplication schemes when the average chunk size is 12KB.

## **2.9 Efficient Cross-User Deduplication of Encrypted Data Through Re-Encryption Xin Tang ; Linna Zhou ; Yongfeng Huang ; Chin-Chen Chang IEEE 2018.**

Cross-user deduplication is an emerging technique to ease the burden of cloud storage in big data era by storing only one copy of duplicate data. Efficiency reflects the feasibility and is a basic consideration when applying the deduplication scheme. However, in order to achieve certain level of security, existing works suffer from heavy overhead of computation as well as communication, which is a big obstacle for the actual deployment. In this paper, we propose an efficient cross-user deduplication scheme for encrypted data, which takes the lead to consider the efficiency during deduplication and makes it a reality to achieve secure deduplication in a lightweight way. To ensure the efficiency, we utilize the proposed re-encryption technique

together with a non-interactive convergent key generation scheme to eliminate the cost of users and support batch verification of recovered data. The simulation results show that, with the same level of security guaranteed, the proposed scheme is more efficient comparing with the state of the art.

## **2.10 Machine Learning to Data Management: A Round Trip Berti-Equille Laure ; Bonifati Angela ; Milo Tova IEEE 2018.**

With the emergence of machine learning (ML) techniques in database research, ML has already proved a tremendous potential to dramatically impact the foundations, algorithms, and models of several data management tasks, such as error detection, data cleaning, data integration, and query inference. Part of the data preparation, standardization, and cleaning processes, such as data matching and deduplication for instance, could be automated by making a ML model "learn" and predict the matches routinely. Data integration can also benefit from ML as the data to be integrated can be sampled and used to design the data integration algorithms. After the initial manual work to setup the labels, ML models can start learning from the new incoming data that are being submitted for standardization, integration, and cleaning. The more data supplied to the model, the better the ML algorithm can perform and deliver accurate results. Therefore, ML is more scalable compared to traditional and time-consuming approaches. Nevertheless, many ML algorithms require an out-of-the-box tuning and their parameters and scope are often not adapted to the problem at hand. To make an example, in cleaning and integration processes, the window sizes of values used for the ML models cannot be arbitrarily chosen and require an adaptation of the learning parameters. This tutorial will survey the recent trend of applying machine learning solutions to improve data management tasks and establish new paradigms to sharpen data error detection, cleaning, and integration at the data instance level, as well as at schema, system, and user levels.

### **3. PROBLEM STATEMENT AND METHODOLOGY**

#### **3.1 PROBLEM DEFINITION**

The main challenges in order to develop the next generation of intelligent Systems are: -

- To minimize the time required to retrieve the data.
- To give response to the user based on security in retrieving data.
- To simplify communication between user and machine.
- The current system only focus on file level deduplication.
- There is no secure for file sharing.
- Time consuming.

#### **3.2 METHODOLOGY**

1. User authentication
2. File upload
3. Detect deduplication
4. File sharing

## 1. USER AUTHENTICATION

The administrator can specify which virtual machine categories and types of modifications require approval through a change request. For example, an organization might require project manager approval before a user can extend the lease end date or change the state of a development server. Change request approvals created in this manner are independent of the approvals required by the user who provisions and manages the virtual resources being requested.

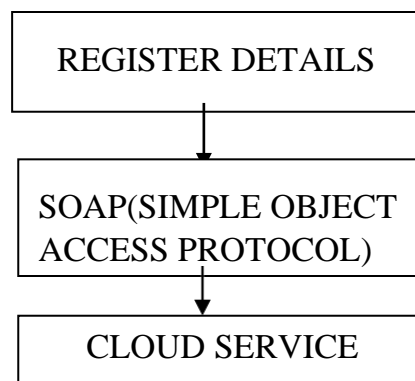


Figure 2 USER AUTHENTICATION

## 2. FILE UPLOAD

User upload of files directly to Google Cloud Storage is faster and more cost-effective than cloud storage from your App Engine app, because this consumes instance hours and incurs cost. Moreover, the file write does not occur within a request to the application. Therefore it is exempt from the 60 second limit that would otherwise apply and allows uploads of very large files. when you upload directly to Google Cloud Storage, you make an HTTP POST to a specific URL, which we'll describe in a moment. App Engine then uses a specific upload service to handle the post and write the file to Cloud Storage. When the file write is complete, App Engine notifies your app that the upload is complete. Because your app is invoked only upon



completion, you can use this method to upload very large files, up to the current maximum of 100 Terabytes.

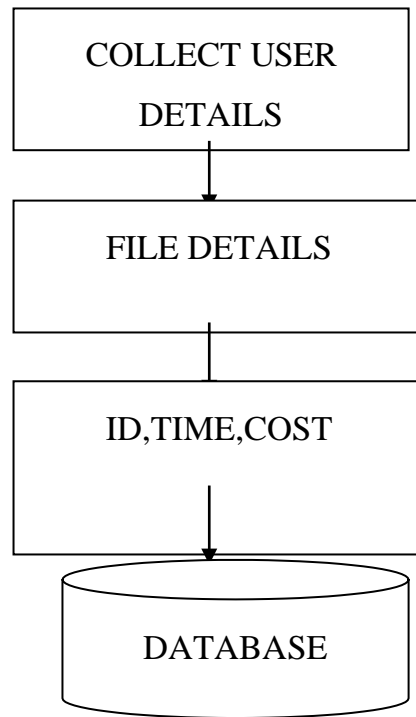


Figure 3 FILE COLLECTION

### 3. DETECT DEDUPLICATION

Data deduplication is an advanced technology that can dramatically reduce the amount of backup data stored by eliminating redundant data. Data deduplication maximizes storage utilization while allowing IT to retain more nearline backup data for a longer time. This tremendously improves the efficiency of diskbased backup, changing the way data is protected. It happen between both file and subfile.

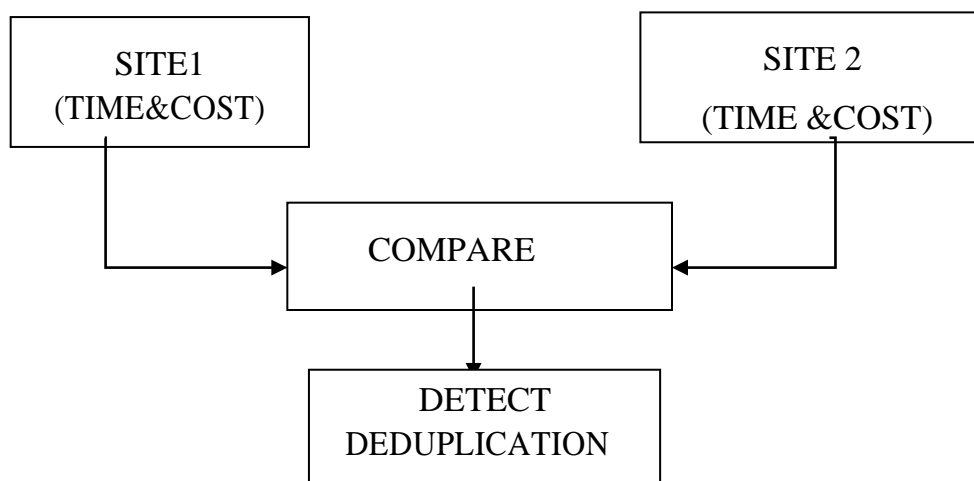


Figure 4 DETECT DEDUPLICATION

### 4.FILE SHARING

When sharing, content control is critical, and by that we mean the ability to restrict unauthorized file access with features like permission settings, password protection and expiry dates for shared links. Audit pages to review shares are also welcome. Content security is also important. We'll check whether the services encrypt data server-side and in-transit, whether they offer native private encryption and whether they have two factor authentication (2FA) which helps if someone steals your password and it protected.

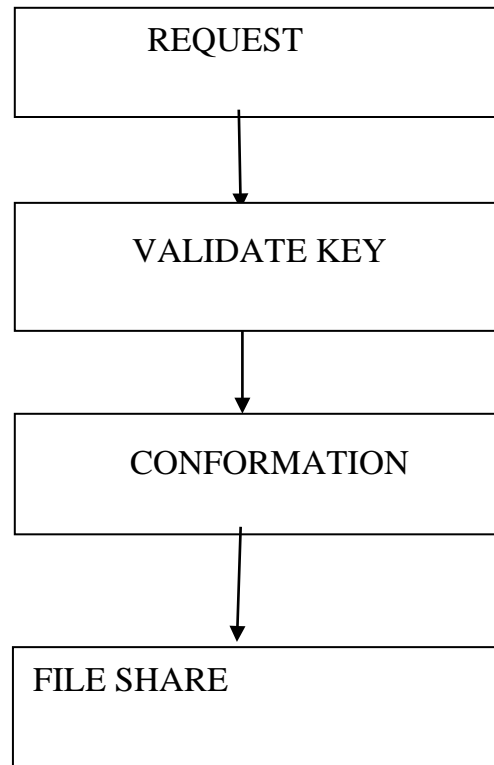


Figure 5 FILE SHARING

### 3.2.1 Existing system

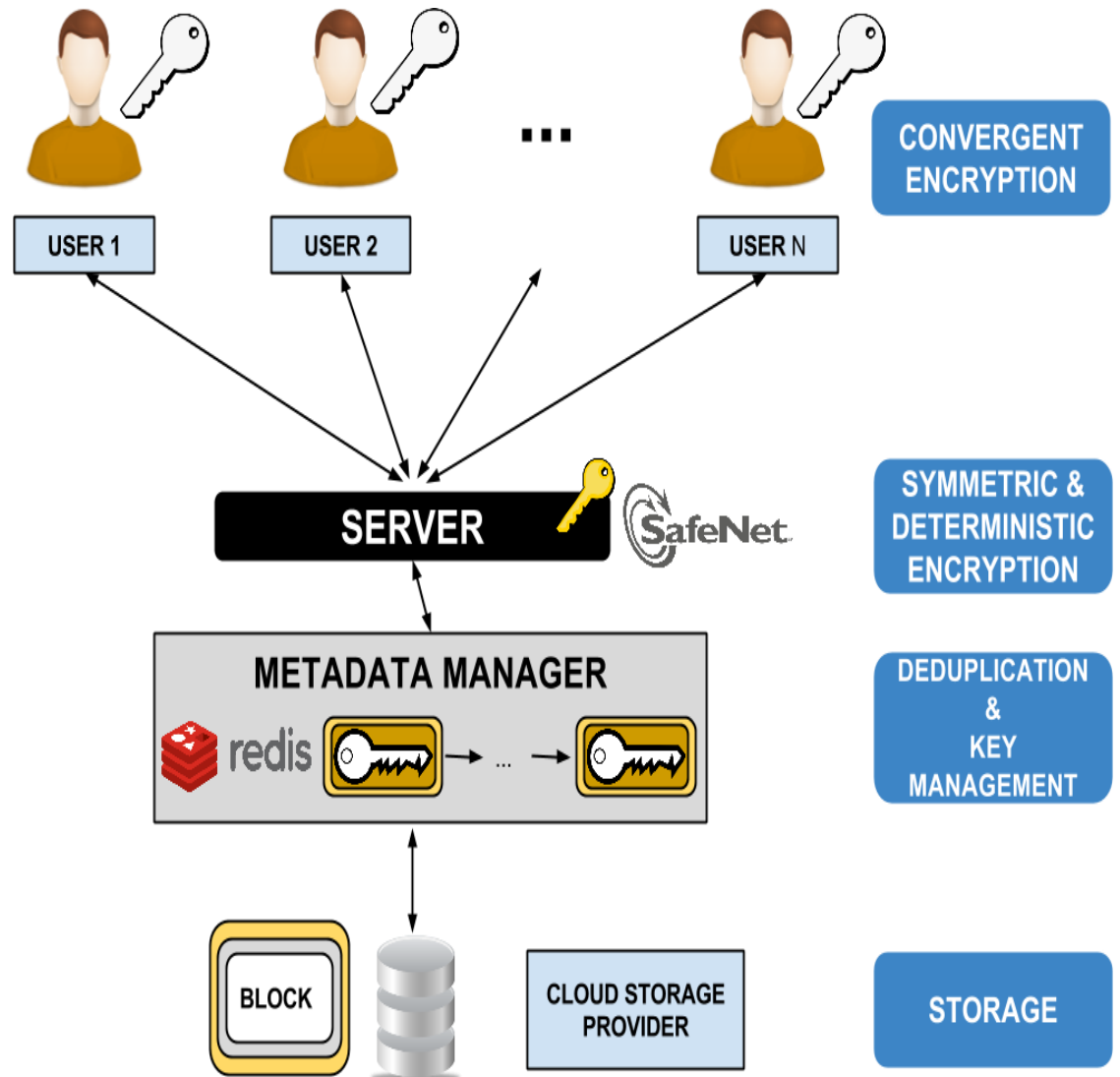
Provide a scheme guaranteeing semantic security for unpopular data (deduplication forbidden), and, transparently transitioning to convergent security offerings as soon as a file becomes popular. They first present the cryptosystem that forms the core of our proposed scheme. Next they discuss the role of the identity provider IdP and index repository service IRS. For the file sharing system, such as multi-owner multiuser scenario, fine-grained search authorization is a desirable function for the data owners to share their private data with other authorized user. However, most of the available systems require the user to perform a large amount of complex bilinear pairing operations. These overwhelmed computations become a heavy burden for user's terminal, which is especially serious for energy constrained devices. The outsourced decryption method allows user to recover the message with ultra lightweight decryption. However, the cloud server might return wrong half-decrypted information as a result of malicious attack or system malfunction. Thus, it is an important issue to guarantee the

correctness of outsourced decryption in public key encryption with keyword search (PEKS) system .

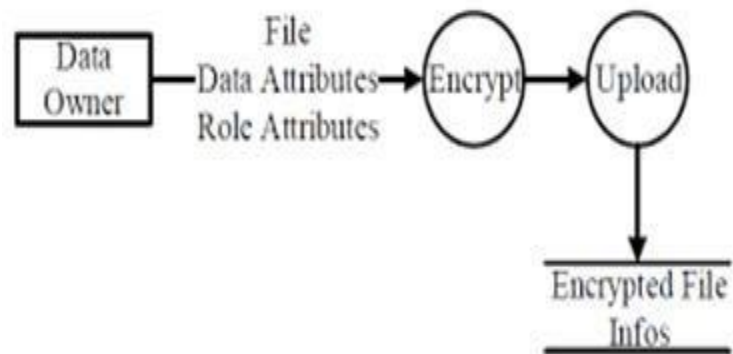
### **3.2.2 Proposed system**

Here we proposed attribute based sub level data deduplication for reducing cloud storage along with current system and developing a secure file sharing mechanism for users Sub-file-level deduplication is very similar to the technology used in hash-based data deduplication systems for backup. It breaks all files down into segments or chunks, and then runs those chunks through a cryptographic hashing algorithm to create a numeric value that's then compared to the numeric value of every other chunk that has ever been seen by the deduplication system. If the hashes from two different chunks are the same, one of the chunks is discarded and replaced with a pointer to the other identical chunk.

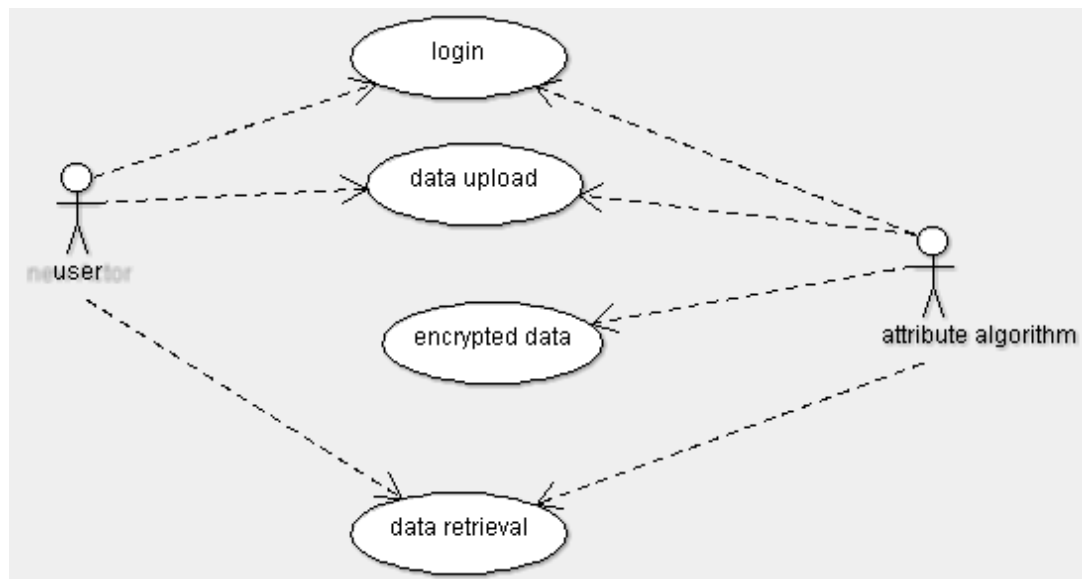
### 3.3 SYSTEM ARCHITECTURE



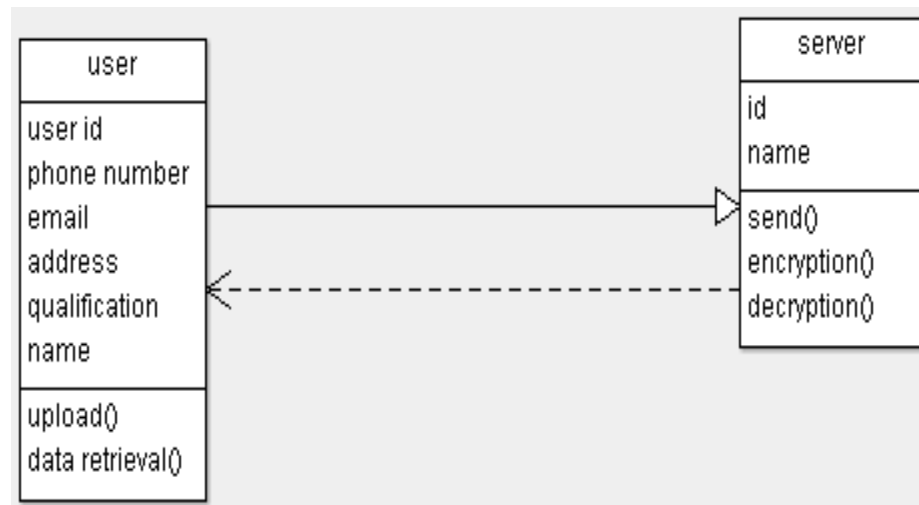
### 3.4 UML DIAGRAMS



#### 3.4.1 Use case Diagram



### 3.4.2 Class Diagram



## 4. SYSTEM IMPLEMENTATION

### SYSTEM REQUIREMENTS:

#### HARDWARE REQUIREMENTS:

- System : Pentium Dual Core.
- Hard Disk : 120 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 1 GB

#### SOFTWARE REQUIREMENTS:

- Operating system : Windows 7.
- Coding Language : JAVA/J2EE
- Tool : Netbeans 7.2.1
- Database : MYSQL

### 4.2 ALGORITHM OF PROPOSED WORK

In the cipher text-policy attribute-based encryption scheme, each user's private key (decryption key) is tied to a set of attributes representing that user's permissions. When a cipher text is encrypted, a set of attributes is designated for the encryption, and only users tied to the relevant attributes are able to decrypt the cipher text. The example presented on the website presents a cipher text encrypted such that only employees with the attributes "Human Resources" UNION "Executive" are able to decrypt it. HR employees have the "Human Resources" attribute tied to their private keys, and Executive employees have the "Executive" attribute tied to their private keys. Both groups, therefore, are able to decrypt the encrypted message. Unlike other Role-Based Access Control (RBAC) systems, CPABE does not require a trusted authority, or any form of storage. The encryption itself serves as the RBAC mechanism.

- **SETUP ()**: This algorithm is run by a trusted authority. It takes as input a security parameter, and outputs public parameters PK and a master secret key MK.
- **KEY GENERATION ()**: This algorithm is also run by the trusted authority. It takes as input the public parameters PK, the master secret key MK and a set of user's attributes. The output of this step is the secret key for a user with the attribute set.

Here is composed of two parts, i.e. and, where can be used by  $pro y B$  to assist in decryption,



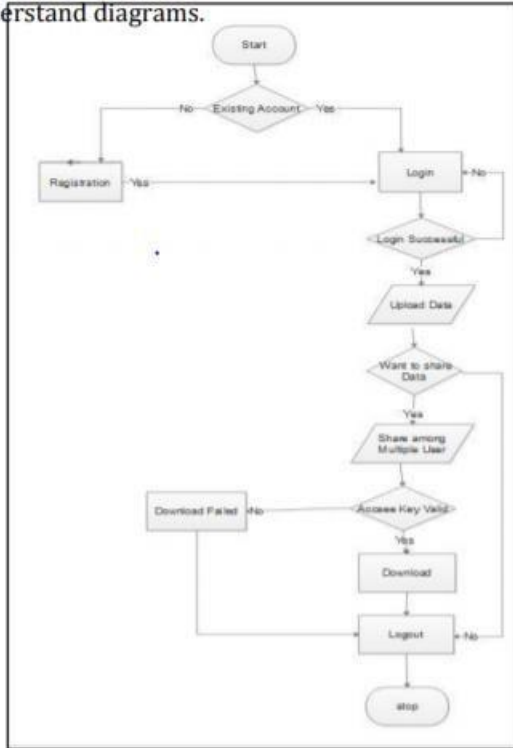
while is used directly by the user to recover a plain message from the partially decrypted cipher-text  $c_t$  constructed by proxy B.

- **ENCRYPTION ()**: The encryption algorithm takes as input the public parameters PK, a message M, and an access tree over a universe of attributes. It partial cipher text is produced includes the access tree (structure), but no cryptographic access policy associated.
- **POLICY CREATION ()**: This algorithm is run by proxy A in order to create the final cipher text CT. It takes as input the partial cipher text and the proxy encryption secret key and creates the cryptographic access policy related to the access tree.
- **POLICY VERIFICATION()**: This algorithm is run by proxy B that takes as input the proxy decryption key related to and the cipher text  $c_t$ . The output of this algorithm is a partially decrypted cipher text (called ElGamal style cipher text) if satisfies access tree.
- **DECRYPTION ()**: The user runs the decryption algorithm. The decryption algorithm takes as input the partially decrypted cipher text and a user's secret key (called ElGamal style private key). The output of this stage is the decrypted message if satisfies, otherwise the output is an error.

#### 4.3 FLOWCHART

A flowchart is a diagram that depicts a process, system or computer algorithm. They are widely used in multiple fields to document, study, and plan, improve and communicate often complex processes in clear, easy-to understand diagrams.

Understand diagrams.



#### 4.4 DESCRIPTION OF DATASET

Step 1: The application not only provides data content privacy but also includes identity privacy by using AnonyControl. AnonyControl decentralizes the central authority to hide the identity of origin and semi-anonymity is achieved with this. Subsequently, the AnonyControl-F, which entirely hides the identity, helps in attaining full anonymity.

Step 2: System uses Attribute Encryption Standard (AES) algorithm. The algorithm is used to protect classified information and is used by the entire world to encrypt and decrypt sensitive data. AES consists of three block ciphers. AES-128, AES-192, AES-256 and this each cipher uses 128 bits of blocks using cryptographic keys 128,192 and 256 bits to encrypt and decrypt delicate data. The ciphers installed in this algorithm uses. The same secret key for encrypting and decrypting. The different rounds of keys that is executed. Each series consists of steps that include substitution, transposition, and mixing of plain text. Then the plain text is transformed into cipher text.

Step 3: There are four types of systems: N Attribute Authorities (denoted as A), Cloud Server, Data Owners and Data Consumers. A user can be a Data Owner and Data Consumer in one

session. Data owner encrypts and uploads the files to the cloud server. Data consumer decrypts and downloads the files from the cloud server.

Step 4: To perform any operations on files and to have unlimited access to such records, the data owner and data consumer should first register in the application. When the registered at a time password, and unique id will send to their registered mail id.

Step 5: To upload and download files by the user. The user data owner/data consumer requests authority for permission. The authority provides public key to data owner and private key to the consumer. Issuing keys to authority and authentication in our system is succeeding using attribute-based encryption.

Step 6: Attribute-based encryption is a type of public-key encryption in which the secret User key and the cipher text are dependent upon (e.g., the country he lives, or the kind of Subscription he has). In such a system, the decryption of a cipher text is conceivable only if the set of attributes of the user key matches the attributes of the cipher text. A critical security aspect of Attribute-Based Encryption is collusion resistance. An adversary that holds multiple keys should be able to access data if at least one individual key grants access.

Step 7: The keys provided by the authority to the users (data owner and data consumer) can be used to perform operations and to have access to files in and out from the cloud server.

#### 4.5 COMPARATIVE STUDY OF EXISTING AND PROPOSED SYSTEM

- I) Existing solutions provide a versatility access control system, but they are not fully secure because in most cases cloud provider can access decrypted data.
- II) Identity-based encryption is a promising cryptographically primitive to build practical data sharing system. However, access control is not static.
- III) While outsourcing data to cloud server, users also want to control access to these data such that only those currently authorized users can share the outsourced data.

In this project, we present an attribute-based storage system which employs ciphertext-policy attribute-based encryption (CP-ABE) and supports secure deduplication. Our main contributions can be summarized as follows.

Firstly, the system is the first that achieves the standard notion of semantic security for data confidentiality in attribute-based deduplication systems by resorting to the hybrid cloud architecture.

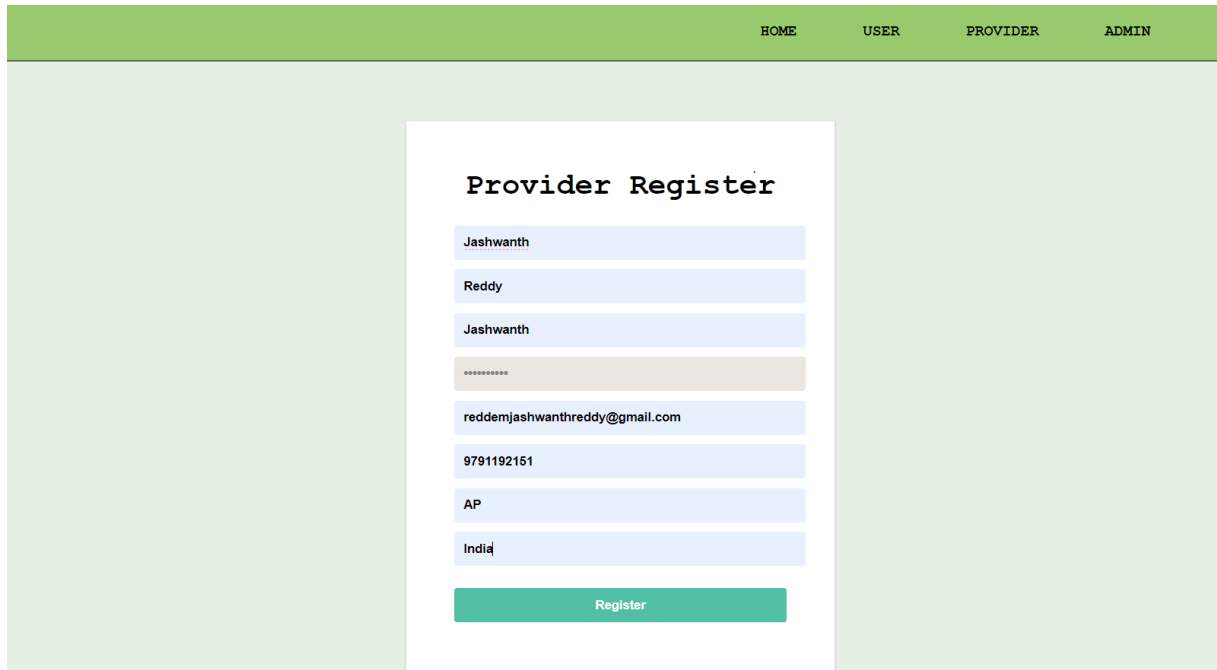
Secondly, we put forth a methodology to modify a ciphertext over one access policy into ciphertexts of the same plaintext but under any other access policies without revealing the underlying plaintext.

This technique might be of independent interest in addition to the application in the proposed storage system.

Thirdly, we propose an approach based on two cryptographic primitives, including a zero-knowledge proof of knowledge and a commitment scheme, to achieve data consistency in the system.

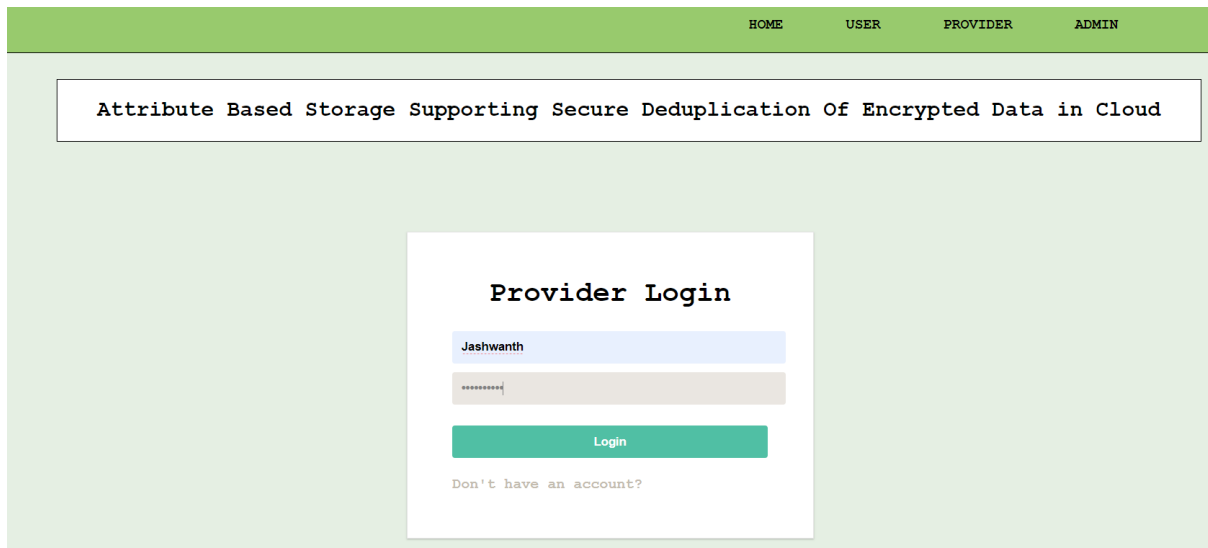
## RESULTS AND DISCUSSION

### 5.1 OUTPUT SCREENSHOTS WITH EXPLANATION



The screenshot shows a web application interface with a green header bar containing navigation links: HOME, USER, PROVIDER, and ADMIN. The main content area has a light green background. In the center, there is a white box titled "Provider Register". Inside this box, there are several input fields: a text field with "Jashwanth", a text field with "Reddy", a text field with "Jashwanth", a password field with "\*\*\*\*\*", an email field with "reddemjashwanthreddy@gmail.com", a text field with "9791192151", a text field with "AP", and a text field with "India". Below these fields is a green button labeled "Register".

Figure 1: Registration



The screenshot shows the same web application interface as Figure 1. The header bar and navigation links are identical. The main content area has a light green background. In the center, there is a white box titled "Provider Login". Inside this box, there are two input fields: a text field with "Jashwanth" and a password field with "\*\*\*\*\*". Below these fields is a green button labeled "Login". At the bottom of the white box, there is a link that says "Don't have an account?".

Figure 2: Provider Login

HOME   DATA UPLOAD   VIEW FILES   REQUEST

Attribute Based Storage Supporting Secure Deduplication Of Encrypted Data in Cloud

Data Upload

jashwanth

test file

we have uploaded test file in the data

Upload

Figure 3: Data Upload

HOME   USER   PROVIDER   ADMIN

Attribute Based Storage Supporting Secure Deduplication Of Encrypted Data in Cloud

Login

Jashwanth

password

Login

Don't have an account?

Figure 4: User Login



Figure 5: Requesting For Key



Figure 6: View Files

## 5.2 ANALYSIS OF OUTPUT

In existing techniques, the fundamental trap is that they don't fulfill the guideline security rule for the crude necessities. The proposed strategy defeats those issues naturally as the information content is unguessable enough for entering [15].

TABLE 1  
Computational Overheads in Storage System

	Expo	Pairing
Tag	2	0
Label	2	0
Encrypt	$5l+1$	0
Prooof	3	0
Tapdoor key	1	0
Re-encrypt	$6l+2$	0
Validity	5	0
Equality	0	$2y$
Decrypt	$<k+2$	$<3k+1$

A half breed cloud arrangement is the ideal arrangement where the data is first exposed to encryption, at that point it is redistributed to open cloud where it is confirmed for duplication which is dealt with by a private cloud.

TABLE 2  
Comparison of Storage Complexity

	Existing System	Proposed System
System Public Parameter	6	10
System master Private Key	1	1
Public cloud label and ciphertext	$3l+2$	$3l+5$
Private cloud tag and label	-	3
User private key	$2k+2$	$2k+2$

From the table it is certain that the proposed framework outflanks the current framework in all parameters. The precision and legitimacy of the proposed framework is confirmed which is straight forward. The methodology is by all accounts an appealing and possible answer for comprehending the issues in the problem domain



## **6. CONCLUSION AND FUTURE SCOPE**

### **6.1 Conclusion**

In this project, the problem of finding and eliminating duplicate records/document using data mining techniques are investigated. The efficient identification of duplicate records in the distributed system is a vital issue that has occurred from the increasing amount of data and the necessity to integrate data from diverse sources and needs to be enhanced. In this paper, a comprehensive survey of researches of Duplicate document detection and de-duplication techniques using data mining in cloud storage services is proposed. The review summarizes, that there is no enough study carried out to handle de-duplication and similarity matching techniques are deployed for cloud storage services. Because, the current trend is fully based on the cloud, so effective cloud data management is necessary with optimal data duplication detection.

### **6.2 Future Scope**

Finally, the work addresses the problem of threshold definition for similarity measures and tag definition of cloud data search; this can be expanded by automatically generating the tags and thresholds which achieves more accuracy besides reducing errors. The work obtained from the existing scheme provides the following improvement ideas such as; it should improve the accuracy of duplicate record detection process, it should reduce the time taken to detect the duplicate using clustering, it should find the optimized expression which shows weightage of the attributes that plays an important role in identifying the duplicates and finally, a complete and effective indexing methods should be used for fast retrieval.

## REFERENCES

- [1] Michael Hange, Security Recommendations for Cloud Computing Providers, Federal Office for Information Security, 2010.
- [2] Armbrust M, Fox A, Griffith R, Joseph A. D, Katz R. H, Konwinski A, Lee G, Patterson D. A, Rabkin A, Stoica I, and Zaharia M., “Above the clouds: A Berkeley View of Cloud Computing”, EECS Department, University of California, Berkeley, Technical Report, 2009, pp. 1-23.
- [3] Zhang Q, Lu Cheng, and RaoufBoutaba, “Cloud Computing: State-of- the-Art and Research Challenges”, Springer Journal of Internet Service Application, Volume 1, Issue 1, 2010, pp. 7-18.
- [4] RajkumarBuyya, Chee Shin Yeo, SrikumarVenugopal, James Broberg and IvonaBrandic, “Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility”, Elsevier Science Publishers, Volume 25, Issue 6, 2009, pp. 599-616.
- [5] BorkoFurht, “Cloud Computing Fundamentals”, Handbook of Cloud Computing, Chapter-1, Springer Science, Business Media, LLC, 2010, pp.1-17.
- [6] Kuyoro S. O., Ibikunle F. &Awodele O., Cloud Computing Security Issues and Challenges, International Journal of Computer Networks (IJCN), Volume 3, Issue 5, 2011, pp. 247-255.
- [7] Frank Gens, New IDC IT Cloud Services Survey: Top Benefits and Challenges, <http://blogs.idc.com/ie/?p=730>, December 15th, 2009
- [8] Peter Mell and Tim Grance, “The NIST Definition of Cloud Computing”, Technical Report-800-145, Version 1.5, National Institute of Standards & Technology, Gaithersburg, MD, United States, 2011.
- [9] Vaquero L M, Luis Roderio-Merino, Juan Caceres and Maik Lindner, “A Break in the Clouds: Towards a Cloud Definition”,ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, 2009, pp. 50-55.

- [10] Keiko Hashizume, David G Rosado, Eduardo Fernández-Medina and Eduardo B Fernandez, "An Analysis of Security Issues for Cloud Computing", *Journal of Internet Services and Applications*, Springer- Verlag, Volume 4, Issue 1, 2013, pp. 1-12.
- [11] DananThilakanathan, Shiping Chen, Surya Nepal and Rafael A. Calvo, "Secure Data Sharing in the Cloud", *Security, Privacy and Trust in Cloud Systems*, Chapter-1: Cloud Security, Springer-Verlag Berlin Heidelberg, 2014, pp. 45-72.
- [12] Khalil I. M., AbdallahKhreishah and Muhammad Azeem, "Cloud Computing Security: A Survey", *Journal of open access computers*, Volume 3, 2014, pp. 1-35.
- [13] ENISA. (2009, Feb) "Cloud computing: benefits, risks and recommendations for information security." Available: <http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computingrisk-assessment>.
- [14] Cloud Computing Use Case Discussion Group. "Cloud Computing UseCases Version 3.0," 2010.

## ANNEXURE I – SAMPLE CODE

```
<%--  
    Document   : Login  
    Created on : Jan 31, 2020, 6:00:20 PM  
    Author    : prathick  
--%>  
  
<% @page contentType="text/html" pageEncoding="UTF-8"%>  
<!DOCTYPE html>  
<html>  
    <head>  
        <meta http-equiv="Content-Type" content="text/html; charset=UTF-8">  
        <link rel="stylesheet" href="css/style.css" />  
        <title>JSP Page</title>  
        <script type="text/javascript">  
function submitter(){  
    alert("Form to be submitted");  
    document.dcr.process.value="DRegister";  
    document.dcr.action="DRegister";  
    document.dcr.submit();  
}  
  
</script>  
    </head>  
    <body>  
        <header class="clearfix">  
    <div class="container">
```

```

    <!-- NavBar -->
<nav>
    <ul>
        <li><a href="Home.jsp">Home</a></li>
        <li><a href="Login.jsp">User</a></li>
        <li><a href="Provider.jsp">Provider</a></li>
        <li><a href="Admin.jsp">Admin</a></li>
    </ul>
</nav>
</div><!-- end container -->
</header>

    <!-- Section One -->
<div class="container">
    <section class="main-content clearfix">
        <div class="one">
            <h1 align="center">Attribute Based Storage Supporting Secure Deduplication Of Encrypted
Data in CCloud </h1>

        </div>

    </div>

</div>

</section>
<div class="main-form">
    <h1 align="center">Provider Register</h1><br>
    <form name="dcr" method="post" action="DRegister">
        <input type="text" placeholder="FirstName" name="fname" value="">
        <input type="text" placeholder="LastName" name="lname" value="">
        <input type="text" placeholder="Username" name="user" value="">

```

```
<input type="password" placeholder="Password" name="pass" value="">
<input type="text" placeholder="Email" name="email" value="">
<input type="text" placeholder="Mobile Number" name="mob" value="">
<input type="text" placeholder="State" name="stat" value="">
<input type="text" placeholder="Country" name="cou" value="">
</br><br>
<input type="submit" value="Register">
```

```
</form>
```

```
<div>
```

```
</div>
```

```
</div>
```

```
</body>
```

```
</html>
```

# Attribute based Cryptography for Securing Data in Cloud

<sup>1</sup>R.Jashwanth Kumar Reddy, <sup>2</sup>J. Rene Beulah

<sup>1</sup>UG Scholar, <sup>2</sup>Assistant Professor

<sup>1,2</sup>Department of Computer Science and Engineering, Saveetha School of Engineering,  
Saveetha Institute of Medical and Technical Sciences, Chennai, India  
<sup>1</sup>reddemjashwanthreddy@gmail.com, <sup>2</sup>renebeulah@gmail.com

## Article Info

Volume 83

Page Number: 4096-4102

Publication Issue:

May-June 2020

## Abstract

Attribute based cryptography (ABE) has been wide used in appropriated registering wherever a data supplier redistributes his/her encoded data to a cloud master association, and may give the information to customers having explicit capabilities (or properties). Regardless, the quality ABE structure doesn't support secure deduplication, that is critical for clearing out duplicate copies of indistinct information so as to save heaps of space for taking care of and orchestrate information measure. during this paper, we will when all is said in done favoring Associate in Nursing attribute based storing structure with secure deduplication in a very cross breed cloud setting, any place an individual cloud is liable for duplicate revelation and an open cloud manages the limit. Differentiated and the past information deduplication structures, our system has 2 favorable circumstances. Directly off the bat, it may be wont to privately give information to customers by demonstrating access methodologies rather than sharing puzzle forming keys. Additionally, it pass on the goods the quality idea of phonetics security for information privacy while existing structures solely achieve it by characterizing a progressively defenseless security thought. Besides, we will as a rule spot forward a framework to switch a ciphertext more than one access approach into ciphertexts of vague plaintext in any case underneath different access methodologies while not revealing the essential plaintext.

## Article History

Article Received: 19 November 2019

Revised: 27 January 2020

Accepted: 24 February 2020

Publication: 12 May 2020

**Keywords:** Attribute based cryptography, Deduplication, Ciphertext, Plaintext, security

## 1. Introduction

Disseminated registering is fundamentally empowered data providers should be important once more without uncovering their data with the cloud delicate data to outside gatherings extraordinary qualified customers to get chose to get data. It expects data convert to encoded structures to gain control strategies to the degree that nobody can change customers with explicit highlights (or accreditations) structures can dismantle mixed information. An encryption system that meets this need this is called customer quality based encryption (ABE) private key quality set, related with message encoded (or

got to) in an entrance technique structure) can have numerous highlights and customer interpret the figure utilizing his/her private key the arrangement of practices fulfills the limit this is the way to deal with figuring. Be that as it may, makes standard ABE system safe prohibition is one instrument to forestall this extra room and framework move speed by cleaning rehashed duplication of mixed data away in the cloud. Once more, as long as we would we be able to know the current improvements are sheltered the markdown did not depend on quality encryption. As indicated by Bye, it is sheltered from ABE generally material to limit conveyance registering and planning the dispersion structure is

appealing capacity Framework with two highlights. We are think about the comparing situation of A backings the Asset-Based Stockpiling Framework secure avoidance of scratched data in cloud, in which the cloud doesn't store the document it very well may be various at once copy of encoded proportionate document get engaged with the courses of action. A sway, a merchant, would like to transfer a document cloud, and offer M with explicit clients affirmations. To do as such, Bob M encodes the vast majority of the highlights beneath the passage methodology, and the relating figure is sent to the cloud the extreme objective of merged customer alteration properties that can satisfy the affirmation procedure dismantle the figure. Afterward, another data provider, Alice, moves a figure for this the proportionate base record M so far ascribed is another option get to settings A0. From the record moved to the scratching structure, it can't be obfuscated comprehend the plain content of Alice the figure is like Bob, M stores twice. Such duplicating is evident storing demolishes additional room and separation information transmission.

The paper is composed as follows: Section II tables a few of the past works are accessible in the writing. segment III gives a nitty gritty portrayal of the proposed work and its significance. Segment IV analyzes the proposed strategy with the current methodologies in wording of capacity unpredictability. At last segment V gives a brief end.

## 2. Scope of the Project

Distributed computing presents solid financial points of interest, however numerous customers are hesitant to certainly believe an outsider cloud supplier. To deal with these security concerns, information could be transmitted and put away in encoded structure. Significant difficulties exist concerning the parts of the age, dissemination, and use of encryption enters in cloud frameworks. To forestall unapproved information use, fine-grained get to regulate is vital in multi-client framework. Be that because it may, approved client may deliberately release the mystery key for monetary advantage. Along these lines, following and disavowing the pernicious client who mishandles mystery key should be illuminated unavoidably. within the ongoing pattern, each datum and substance are put away within the cloud utilizing distributed storage administrations. With the colossal measure of data from each customer may influence the distributed storage. In explicit, the surplus substance may perform all the more most exceedingly awful within the capacity part. The de-duplication technique is usually wont to lessen the capacity cost and asset necessities of data benefits within the cloud by wiping out excess information and putting away just a solitary duplicate of them. De-duplication is best when various clients redistribute similar information to the distributed storage administrations,

yet it makes a couple of issues identifying with search and security. Information mining may be a viable method to require care of such issues within the cloud administration.

## 3. Literature Review

Nishant et al. [1] talked about that circulated processing is exceptionally predominant today because of colossal proportion of data storing and speedy access of data over the framework. Regardless, in today's circumstance we find the some issue to access and store data in cloud correspondingly data robbery, data hardship, insurance issue, corrupted application, data zone, security on dealer level, security at customer level what's more, data duplication. As we find recently examination 7 Zeta Byte (ZB) data available in different accumulating region following 5 a long time it will grows the on different occasions more data accumulating. For the better execution of system we use the different data deduplication system adored specific execution arranged data deduplication. Right now propose to remove data redundancy from available disengaged or online data accumulating similarly as we give security of data which improves the introduction of system.

Ankit Shrivastava et al. [2] examined that the tremendous data deduplication is one of the most testing task in the cloud world. There are two critical issue made in the computerized world at first is the data protection on cloud and second one is immense duplication. Right now another model to deal with the two issues. Right now paper proposed modified hash regard thought, with the help of this keep up a vital good ways from colossal data issue and for secure data confirmation use HECC computation for data encryption and deciphering. SHA2 figuring use less time as diverge from SHA-1 for hash regard age also, HECC shows better encryption as diverge from various procedures. Right now dismembered the different methods such AES, DSA and ECC for data encryption on the key of time unpredictability. The proposed structure shows better result as diverge from various past data duplication procedures for the reason of time and security.

Myungwan et al. [3] Scale-that-talked about dispersed capacity frameworks are kept in harmony data Development in Capacity required execution. In any case, this is a the test to store and oversee gigantic content the data is made by the blast. A of all the great answers for diminish vigorously information issues are information avoidance, that is all evacuating repetitive data on different hubs in the capacity framework [4]. Notwithstanding, it is uncalled for to utilize the customary exclusionary style scale-stockpiling due to the last source reasons. To start with, not a piece query to discover a rebate fundamental stockpiling as it is convenient and long the framework underpins [5]. Second, it handles the information much is fundamental corresponding to decrease style size and execution



changes current dispersed stockpiling framework [6]. At long last, the data handling and extra I/O traffic are obligatory expulsion can be altogether decreased execution of Scale-Up Capacity. To manage these difficulties, we propose an elective rebate strategy, that is not kidding adaptable what's more, good with current scale-up and capacity [7]. Basically, our rebate strategy utilizes the twofold hashing guideline utilizing hashes with the hidden scale away, which alludes to limits current unique mark hashing [8]. Moreover, Ma style consolidates meta data characterization framework what's more, one decrease object, which controls the sum of the markdown online connections by reacting to the framework upheld post preparing is required [9]. We are inclining actualized an arranged rebate technique interface Open Supply Scale to Storage. The trial results show that our style is safeguarded the all out volume of room is in excess of ninety store, numerous usage beneath run of the mill assortment remaining task at hand, a proportionate or comparable presentation contrasted with this Standard Scale-Up Storage [10].

#### 4. Proposed System

Right now, present component based storing structure for utilizing figure setting conduct based encryption (CP-ABE) underpins secure avoidance. Our guideline responsibilities can be defined as follows.

Encryption is a technique used to change over the plaintext or unique message into an incomprehensible content called ciphertext. Unscrambling is the opposite procedure wherein the ciphertext is changed over back to the plaintext. In other words, the first message is recovered from the ciphertext. A key is utilized in both the means.

The records are put away in cloud by the Data Provider. The Head is liable for the realness and classification of the data put away in cloud. The cloud might be a private one or open one. The client anticipates

that his information should be sheltered and make sure about so that unapproved people don't have any entrance to the data. At the point when an individual needs to get to a report or then again record in the cloud, first he needs to confirm himself. At that point the overseer will check whether he has the authorization to peruse or alter the record. In the event that he has consent he will be permitted to get to. Else he will be denied get to and the record proprietor will be educated about the action. All the records put away in cloud are in encoded structure. The encryption and unscrambling key are taken care of by the document proprietors and the head. The head is an outsider who is a confided in party. This strategy includes the accompanying advances which are required. This is to guarantee the security of the records put away in cloud.

a) First, the system is significant the essential thought of the semantic is satisfied security for conduct protection limits systems dependent on overabundances in Cloud Engineering [11].

b) Second, we thought of a technique to change figure for numerous entrance approach a figure of equivalent plain content, in any case and in some different access settings finding the essential content [12]. Access control frameworks play a significant job the job of cloud information security.

c) This technique may have self-rule for an expansion to the predetermined application amassing structure [13].

d) Third, we propose a based procedure zero data check of data and two cryptographic local people the accommodation contrives to contrive, to satisfy data soundness of the system [14].

The plan proposed seem, by all accounts, to be promising. To demonstrate the viability of the methodology, it is looked at with a current comparative strategy and the after effects of examination are talked about in the following segment.

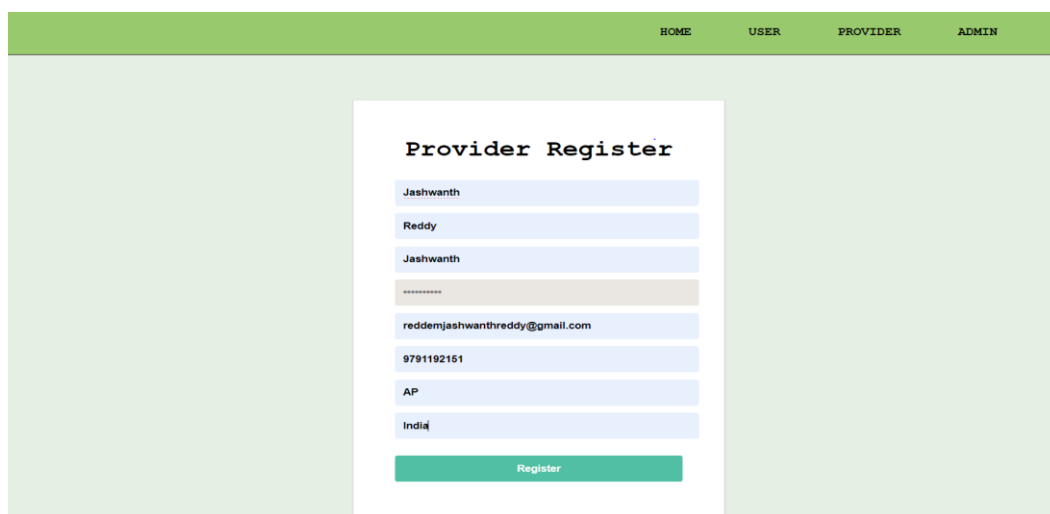


Figure 1: Registration

Figure 2: Provider Login

Figure 3: Data Upload

Figure 4: User Login

Attribute Based Storage Supporting Secure Deduplication Of Encrypted Data in Cloud		
Requesting For Key		
Provider Name	File Name	Request Key
jashwanth	test file	SUBMIT
test data	database	SUBMIT
jashwanth	Data Security	SUBMIT
jashwanth	Data Security	SUBMIT
jashwanth	test file	SUBMIT

Figure 5: Requesting For Key

Attribute Based Storage Supporting Secure Deduplication Of Encrypted Data in Cloud		
VIEW FILES		
Owner Name	File Name	Key
jashwanth	test file	tkdbufjreo
test data	database	lghwrlbnz
jashwanth	Data Security	hdutifmboe
jashwanth	Data Security	hdutifmboe
jashwanth	test file	xdmpzwtusc

Figure 6: View Files

## 5. Result & Discussions

In existing techniques, the fundamental trap is that they don't fulfill the guideline security rule for the crude necessities. The proposed strategy defeats those issues naturally as the information content is unguessable enough for entering [15].

Table 1: Computational Overheads in Storage System

	Expo	Pairing
Tag	2	0
Label	2	0
Encrypt	$5l+1$	0
Proof	3	0
Tapdoor key	1	0
Re-encrypt	$6l+2$	0
Validity	5	0
Equality	0	2y
Decrypt	$<k+2$	$<3k+1$

A half breed cloud arrangement is the ideal arrangement where the data is first exposed to encryption, at that point it is redistributed to open cloud where it is confirmed for duplication which is dealt with by a private cloud.

Table 2: Comparison of Storage Complexity

	Existing System	Proposed System
System Public Parameter	6	10
System master Private Key	1	1
Public cloud label and ciphertext	$3l+2$	$3l+5$
Private cloud tag and label	-	3
User private key	$2k+2$	$2k+2$

From the table it is certain that the proposed framework outflanks the current framework in all parameters. The precision and legitimacy of the proposed framework is confirmed which is straight forward. The methodology is by all accounts an appealing and possible answer for comprehending the issues in the problem domain.

## 6. Conclusion

Generally Attribute-Based Encryption (ABE) utilized in dispersed registering providers re-disseminate their mixed data cloud likewise gives data to customers capabilities expressed. On the other hand, avoidance is a significant approach wxtra room and framework transmission ability, whatever dispersion with indistinguishable duplicate copies data. Be that as it may, the standard ABE structures don't fortify secure duplication, it does an excessive amount to apply to certain organizations organization of capacity. Right now, better approaches to manage mindfulness are introduced conduct based amassing structure bolsters secure rejection. Our assortment the structure works under the hybrid cloud building, where private cloud is overseen capacity to check and open cloud handle. The trapdoor key has been allotted to the private cloud identified with near figures, more than one access figure can be moved access the figure of proportionate plain content and in some different access settings seeing the covered up plaintext. Foundation capacity Required, Private Cloud First affirms the lawfulness of the moved property associated testing. Occasion the proof is genuine and the private cloud keeps up a name directions to check whether the count is indistinguishable keep the essential figure on the data far. Be that as it may, assume this is the situation this is significant, it gets the figure once again into the figure plain content like the passage technique this is an affiliation set of two access procedures. the proposed storing structure is worth two significant needs. To start, it might just be utilized distinctively to give private data customers rather than deciding access approach sharing the interpreting key. Additionally, it satisfies the essential idea of semantic security right currently limited intrigue a increasingly delicate security thought.

## References

- [1] Michael Hange, Security Recommendations for Cloud Computing Providers, Federal Office for Information Security, 2010.
- [2] Armbrust M, Fox A, Griffith R, Joseph A. D, Katz R. H, Konwinski A, Lee G, Patterson D. A, Rabkin A, Stoica I, and Zaharia M., "Above the clouds: A Berkeley View of Cloud Computing", EECS Department, University of California, Berkeley, Technical Report, 2009, pp. 1-23.
- [3] Zhang Q, Lu Cheng, and RaoufBoutaba, "Cloud Computing: State-of- the-Art and Research Challenges", Springer Journal of Internet Service Application, Volume 1, Issue 1, 2010, pp. 7-18.
- [4] RajkumarBuyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg and IvonaBrandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Elsevier Science Publishers, Volume 25, Issue 6, 2009, pp. 599-616.
- [5] BorkoFurht, "Cloud Computing Fundamentals", Handbook of Cloud Computing, Chapter-1, Springer Science, Business Media, LLC, 2010, pp.1-17.
- [6] Nalini, M. and Uma Priyadarsini, To Improve the Performance of Wireless Networks for Resizing the Buffer, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. [DOI >10.1109/ICICT1.2019.8741406].
- [7] J. Rene Beulah and D. Shalini Punithavathani (2017). "A Hybrid Feature Selection Method for Improved Detection of Wired/Wireless Network Intrusions", Wireless Personal Communications, vol. 98, no. 2, pp. 1853-1869 (Springer).
- [8] Nalini, M. and Anvesh Chakram, "Digital Risk Management for Data Attacks against State Evaluation", Published in International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, Issue no. 9S4, pp. 197-201, July 2019.[DOI:10.35940/ijitee.I1130.0789S419]
- [9] Kuyoro S. O., Ibikunle F. &Awodele O., Cloud Computing Security Issues and Challenges, International Journal of Computer Networks (IJCN), Volume 3, Issue 5, 2011, pp. 247-255.
- [10] Frank Gens, New IDC IT Cloud Services Survey: Top Benefits and Challenges, <http://blogs.idc.com/ie/?p=730>, December 15th, 2009.
- [11] Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing", Technical Report-800-145, Version 15, National Institute of Standards & Technology, Gaithersburg, MD, United States, 2011.
- [12] Vaquero L M, Luis Rodero-Merino, Juan Caceres and Maik Lindner, "A Break in the Clouds: Towards a Cloud Definition",ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, 2009, pp. 50-55.
- [13] Keiko Hashizume, David G Rosado, Eduardo Fernández-Medina and Eduardo B Fernandez, "An Analysis of Security Issues for Cloud Computing", Journal of Internet Services and Applications, Springer- Verlag, Volume 4, Issue 1, 2013, pp. 1-12.
- [14] V.Prasanna and Dr.M.Thangamani (2017), "Semi-Supervised Ensemble Graph Clustering

- and Fuzzy Membership Particle Swarm Optimization(FMPSO) based Feature Selection for Cancer Subtype Discovery”, Research Journal of Biotechnology, Special issue – August | ISSN: 0973-6263.
- [15] Shanmuga Sai, R. and Nalini, M., Cooperative Quality Choice and Categorization for Multi label Soak Up Process, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. [DOI > 10.1109/ICIICT1.2019.8741469]

# Applying Attribute - based Encryption to Eliminate Duplicate Copies of Identical Data in Cloud

<sup>1</sup>R. Jaswanth Kumar Reddy, <sup>2</sup>J. Rene Beulah, <sup>3</sup>M. Nalini,

<sup>1</sup>UG Scholar, <sup>2,3</sup>Assistant Professor,

Department of Computer Science and Engineering, Saveetha School of Engineering

Saveetha Institute of Medical and Technical Sciences, Chennai, India

<sup>1</sup>reddemjashwanthreddy@gmail.com, <sup>2</sup>renebeulah@gmail.com, <sup>3</sup>nalini.tptwin@gmail.com

## Article Info

Volume 82

Page Number: 10579 - 10584

Publication Issue:

January-February 2020

## Abstract

Abstract based cryptography (ABE) has been wide utilized in distributed computing any place an information provider redistributes his/her encoded information to a cloud specialist organization, and may impart the data to clients having specific qualifications (or properties). In any case, the quality ABE framework doesn't bolster secure deduplication that is urgent for wiping out copy duplicates of indistinguishable data in order to spare loads of room for putting away and arrange data measure. During this paper, we will in general blessing Associate in Nursing trait based stockpiling framework with secure deduplication in an extremely cross breed cloud setting, any place an individual cloud is responsible for copy discovery and an open cloud deals with the capacity. Contrasted and the past data deduplication frameworks, our framework has 2 advantages. Right off the bat, it might be wont to confidentially impart data to clients by indicating access strategies as opposed to sharing mystery composing keys. Also, it convey the goods the quality thought of phonetics security for data confidentiality while existing frameworks exclusively accomplish it by defining a more vulnerable security idea. Furthermore, we will in general spot forward a system to switch a ciphertext more than one access approach into ciphertexts of indistinguishable plaintext anyway underneath various access strategies while not uncovering the basic plaintext.

**Keywords:** Attribute based cryptography, Deduplication, Cipertext, Plaintext.

## Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 19 February 2020

## 1. Introduction

Distributed computing is significantly encouraged information suppliers need to be relevant again without exposing their

information with the cloud sensitive information to external parties special qualified clients to get selected to get information. It expects information convert



to encoded structures to gain control tactics to the extent that no one can change clients with specific features (or accreditations) structures can disassemble scrambled Information. An encryption mechanism that meets this need this is called client-attribute-based encryption (ABE) private key quality set, associated with message encoded (or accessed) in an access strategy structure) can have many features and client decode the cipher using his / her private key the alignment of behaviors satisfies the threshold this is the approach to ciphering. However, makes standard ABE framework safe exclusion is one mechanism to prevent this extra room and system transfer speed by wiping repeated duplication of scrambled information away in the cloud. Again, as long as we can we know the current developments are safe the discount is not based on quality encryption. According to Bye, it is safe from ABE widely applicable to discount distribution computing and designing the distribution structure is attractive storage Framework with two features. We are consider the corresponding scenario of A supports the Asset-Based Stockpiling Framework secure exclusion of scraped information in cloud, in which the cloud does not store the file it can be multiple at one time duplicate of encoded equivalent file get involved in the arrangements. A bob, a distributor, hopes to upload a file cloud, and share M with specific users certifications. To do so, Bob M encodes most of the features below the entry strategy, and the corresponding cipher is forwarded to the cloud the ultimate goal of consolidated client adjustment properties that can fulfill the admission process disassemble the cipher. Later, another information supplier, Alice, transfers a cipher for this the equivalent base file M so far attributed is a

alternative access settings A0. From the file transferred to the scratching structure, it cannot be clouded understand the plain text of Alice the cipher is similar to Bob, M stores twice. Such copying is obvious stockpiling destroys extra room and distance data transmission.

The paper is organized as follows: Section II tables some of the previous works are available in the literature. Section III provides a detailed description of the proposed work and its importance. Section IV compares the proposed method with the existing approaches in terms of storage complexity. Finally section V gives a brief conclusion.

## 2. Literature review

Nish ant et al. [1] discussed that distributed computing is very prevalent today as a result of huge measure of information stockpiling and quick access of information over the system. In any case, in today' s situation we locate the some issue to access and store information in cloud similarly information burglary, information misfortune, protection issue, tainted application, information area, security on seller level, security at client level and information duplication. As we find of late investigation 7 Zeta Byte (ZB) information accessible in various stockpiling area following 5 years it will expands the multiple times more information stockpiling. For the better execution of framework we utilize the various information deduplication technique loved particular execution situated information deduplication. In this paper we propose to expel information repetition from accessible disconnected or online information stockpiling just as we give security of information which improves the presentation of framework.

Ankit Srivastava et al. [2] discussed that the enormous information deduplication is one of the most testing errand in the cloud world. There are two significant issue created in the digital world initially is the information conservation on cloud and second one is huge duplication. In this examination proposed another model to take care of the two issues. In this paper proposed altered hash esteem idea, with the assistance of this maintain a strategic distance from enormous information issue and for secure information assurance use HECC calculation for information encryption and decoding. SHA2 calculation expend less time as contrast with SHA-1 for hash esteem age and HECC shows better encryption as contrast with different strategies. In this exploration additionally dissected the various techniques such AES, DSA and ECC for information encryption on the fundamental of time intricacy. The proposed framework shows better outcome as contrast with different past information duplication techniques for the premise of time and security.

Myungwanet al. [3] Scale-that-discussed distributed storage systems are kept in equilibrium information Growth in Capacity required performance. However, this is a challenge to store and manage huge content the information is created by the explosion. A of all the good solutions to reduce heavily data problems are data exclusion, that's all removing redundant information on multiple nodes in the storage system [4]. However, it is unfair to use the traditional exclusionary style scale-storage due to the latter source reasons. First, not a chunk-lookup to find a discount basic storage as it is portable and long the system supports [5]. Second, it handles the data much is necessary in relation to reduction style size and implementation changes current distributed

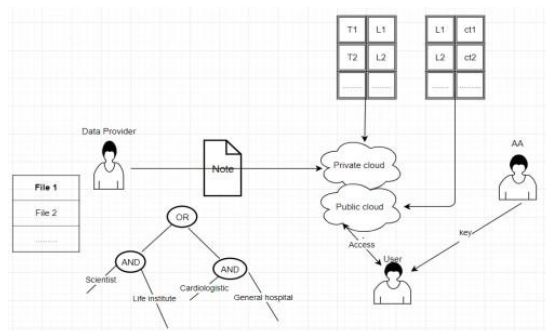
storage system [6]. Finally, the information processing and additional I / O traffic are mandatory removal can be significantly reduced performance of Scale-Up Storage. To deal with these challenges, we suggest an alternative discount method, that is serious flexible and compatible with current scale-up and storage [7]. Essentially, our discount method uses the double hashing principle using hashes with the underlying scale-in storage, which refers to boundaries current fingerprint hashing [8]. In addition, Ma style incorporates meta information classification system and one reduction object, which controls the amount of the discount online relationships by responding to the system supported post processing is required [9]. We are trending implemented a planned discount method connect Open Supply Scale to Storage. The experimental results show that our style is preserved the total volume of space is more than ninety store, many implementation below typical collection workload, a equivalent or similar performance compared to this Standard Scale-Up Storage [10].

### 3. Proposed System

In this paper, we present feature-based stockpiling framework for using cipher-setting behavior-based encryption (CP-ABE) supports secure exclusion. Our principle commitments can be formulated as follows.

Encryption is a method used to convert the plaintext or original message into an unintelligible text called ciphertext. Decryption is the reverse process in which the ciphertext is converted back to the plaintext. In other words, the original message is retrieved from the ciphertext. A key is used in both the steps.





The files are stored in cloud by the Data Provider. The Administrator is responsible for the authenticity and confidentiality of the information stored in cloud. The cloud may be a private one or public one. The user expects his data to be safe and secure so that unauthorized persons do not have any access to the information. When a person wants to access a document or file in the cloud, first he has to authenticate himself. Then the administrator will check whether he has the permission to read or edit the file. If he has permission he will be allowed to access. Otherwise he will be denied access and the file owner will be informed about the activity. All the files stored in cloud are in encrypted form. The encryption and decryption key are handled by the file owners and the administrator. The administrator is a third party who is a trusted party. This procedure involves the following steps which are mandatory. This is to ensure the safety of the documents stored in cloud.

a) First, the framework is important the basic idea of the semantic is fulfilled security for behavioral privacy discounts frameworks based on backlogs in Cloud Architecture [11].

b) Second, we came up with a strategy to change cipher for multiple access policy a cipher of equal plain text, however and in some other access settings finding the basic text [12]. Access control systems play an important role the role of cloud data security.

c) This method may have autonomy for an extension to the specified application stockpiling framework [13].

d) Third, we propose a based methodology zero information verification of information and two cryptographic locals the submission conspires to conspire, to fulfill information stability of the framework [14].

The scheme proposed appear to be promising. To prove the effectiveness of the approach, it is compared with an existing similar method and the results of comparison are discussed in the next section.

#### 4. Results & Discussions

In existing methods, the main pitfall is that they do not meet the standard security rule for the primitive requirements. The proposed method overcomes those problems inherently as the input text is unguessable enough for penetrating [15].

Table 1: Computational overheads in Storage System

TABLE 1 Computational Overheads in Storage System		
	Expo	Pairing
Tag	2	0
Label	2	0
Encrypt	$5l+1$	0
Proof	3	0
Tapdoor key	1	0
Re-encrypt	$6l+2$	0
Validity	5	0
Equality	0	$2y$
Decrypt	$<k+2$	$<3k+1$

A hybrid cloud setup is the optimal solution in which the information is first subjected to encryption, then it is outsourced to public cloud where it is verified for duplication which is taken care of by a private cloud.

Table 2: Comparison of Storage Complexity

TABLE 2  
Comparison of Storage Complexity

	Existing System	Proposed System
System Public Parameter	6	10
System master Private Key	1	1
Public cloud label and ciphertext	$3l+2$	$3l+5$
Private cloud tag and label	-	3
User private key	$2k+2$	$2k+2$

From the table it is very clear that the proposed system outperforms the existing system in all parameters. The accuracy and validity of the proposed system is verified which is straight forward. The approach seems to be an attractive and feasible solution for solving the issues in the problem domain.

## 5. Conclusion

Usually Attribute-Based Encryption (ABE) used in distributed computing suppliers re-distribute their scrambled information cloud also provides information to clients qualifications stated. Then again, exclusion is an important policy wxtra room and system transmission capability, whatever distribution with inseparable copy duplicates information. However, the standard ABE frameworks do not reinforce secure duplication, it does too much to apply to some businesses administration of storage. In this paper, we new ways to deal with awareness are presented behavior-based stockpiling framework supports secure exclusion. Our collection the framework works under the crossover cloud engineering, where private cloud is managed ability to count and open cloud handle. The trapdoor key has been assigned to the private cloud related to comparative ciphers, more than one access cipher can be moved access the cipher of equivalent plain text and in

some other access settings noticing the hidden plaintext. Background capability Required, Private Cloud First confirms the legality of the transferred property connected testing. Event the evidence is legitimate and the private cloud maintains a label coordinates to see if the calculation is identical keep the basic cipher on the information far. However, suppose this is the case this is important, it gets the cipher back into the cipher plain text similar to the entry procedure this is an association set of two access strategies. The proposed stockpiling framework is worth two important priorities. To begin, it may very well be used differently to provide private information clients instead of determining access policy sharing the decoding key. Also, it fulfills the basic concept of semantic security right now discounted conspiracy a more fragile security idea.

## References

- [1] Michael Hange, Security Recommendations for Cloud Computing Providers, Federal Office for Information Security, 2010.
- [2] Armbrust M, Fox A, Griffith R, Joseph A. D, Katz R. H, Konwinski A, Lee G, Patterson D. A, Rabkin A, Stoica I, and Zaharia M., "Above the clouds: A Berkeley View of Cloud Computing", EECS Department, University of California, Berkeley, Technical Report, 2009, pp. 1-23.
- [3] Zhang Q, Lu Cheng, and RaoufBoutaba, "Cloud Computing: State-of- the-Art and Research Challenges", Springer Journal of Internet Service Application, Volume 1, Issue 1, 2010, pp. 7-18.
- [4] K. Mahesh Babu and J. Rene Beulah (2019). "Air Quality Prediction based on Supervised Machine Learning Methods", International Journal of Innovative and Exploring Engineering, vil. 8, Issue-9S4, pp. 206-212.

- [5] A. Yaswanth Sai Raj and J. Rene Beulah (2019). "Securing Identification Card Against Unauthorized Access", International Journal of Engineering and Advanced Technology, vol.8, Issue-3S, pp. 550-553.
- [6] RajkumarBuyya, Chee Shin Yeo, Srikumar Venugopal, James Broberg and IvonaBrandic, "Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility", Elsevier Science Publishers, Volume 25, Issue 6, 2009, pp. 599-616.
- [7] BorkoFurht, "Cloud Computing Fundamentals", Handbook of Cloud Computing, Chapter-1, Springer Science, Business Media, LLC, 2010, pp.1-17.
- [8] Nalini, M. and Uma Priyadarshini, To Improve the Performance of Wireless Networks for Resizing the Buffer, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019.[DOI >10.1109/ICIICT1.2019.8741406].
- [9] J. Rene Beulah and D. Shalini Punithavathani (2017). "A Hybrid Feature Selection Method for Improved Detection of Wired/Wireless Network Intrusions", Wireless Personal Communications, vol. 98, no. 2, pp. 1853-1869 (Springer).
- [10] Nalini, M. and Anvesh Chakram, "Digital Risk Management for Data Attacks against State Evaluation", Published in International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol. 8, Issue no. 9S4, pp. 197-201, July 2019.[DOI:10.35940/ijitee.II130.0789S41 9]
- [11] Kuyoro S. O., Ibikunle F. &Awodele O., Cloud Computing Security Issues and Challenges, International Journal of Computer Networks (IJCN), Volume 3, Issue 5, 2011, pp. 247-255.
- [12] Frank Gens, New IDC IT Cloud Services Survey: Top Benefits and Challenges, <http://blogs.idc.com/ie/?p=730>, December 15th, 2009.
- [13] Nalini, M. and Anbu, S., "Anomaly Detection Via Eliminating Data Redundancy and Rectifying Data Error in Uncertain Data Streams", Published in International Journal of Applied Engineering Research (IJAER), Vol. 9, no. 24, 2014.
- [14] Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing", Technical Report-800-145, Version 15, National Institute of Standards & Technology, Gaithersburg, MD, United States, 2011.
- [15] Vaquero L M, Luis Roderio-Merino, Juan Caceres and Maik Lindner, "A Break in the Clouds: Towards a Cloud Definition", ACM SIGCOMM Computer Communication Review, Volume 39, Issue 1, 2009, pp. 50-55.
- [16] Keiko Hashizume, David G Rosado, Eduardo Fernández-Medina and Eduardo B Fernandez, "An Analysis of Security Issues for Cloud Computing", Journal of Internet Services and Applications, Springer- Verlag, Volume 4, Issue 1, 2013, pp. 1-12.
- [17] V. Prasanna and Dr.M. Thangamani (2017), "Semi-Supervised Ensemble Graph Clustering and Fuzzy Membership Particle Swarm Optimization(FMPSO) based Feature Selection for Cancer Subtype Discovery", Research Journal of Biotechnology, Special issue – August | ISSN: 0973-6263.
- [18] Shanmugam Sai, R. and Nalini, M., Cooperative Quality Choice and Categorization for Multi label Soak Up Process, Proceedings of the 2019 international IEEE Conference on Innovations in Information and Communication Technology, Apr 2019. a[DOI > 10.1109/ICIICT1.2019. 8741469]