A PROJECT REPORT
ON

# Smart Search Over Enciphered Data In Cloud Computing

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

**BACHELOR OF ENGINEERING**

Submitted by

| | |
|---|---|
| **Gauri Kavitkar** | **B151024252** |
| **Sarvesh Kher** | **B151024330** |
| **Gaurav Mahendrakar** | **B151024334** |
| **Mayuresh Patil** | **B151024336** |

Under the guidance of

**Ms.Nikhita Nerkar**

DEPARTMENT OF COMPUTER ENGINEERING

**RMD Sinhgad School of Engineering**
**Warje, Pune-58**

**SAVITRIBAI PHULE PUNE UNIVERSITY**

**2018 -2019**

Sinhgad Institutes

# CERTIFICATE

This is to certify that project entitled

## Smart Search Over Enciphered Data In Cloud Computing

Submitted by

| | |
|---|---|
| **Gauri Kavitkar** | **B151024252** |
| **Sarvesh Kher** | **B151024330** |
| **Gaurav Mahendrakar** | **B151024334** |
| **Mayuresh Patil** | **B151024336** |

is a bonafide work carried out by Students under the supervision of Ms.Nikhita Nerkarand it is submitted towards the partial fulfillment of the requirement of Bachelor of Engineering (Computer Engineering).

**Ms.Nikhita Nerkar**
Guide,
Department of Computer Engineering

**Prof.Parth Sagar**
Project Co-ordinator,
Department of Computer Engineering

**Prof. V. M.Lomte**
Head,
Department of Computer Engineering

**Dr. V.V.Dixit**
Principal,
RMDSSOE, Warje, Pune 58

Place: Pune
Date: / / 2019

# ACKNOWLEDGEMENT

I have great pleasure in presenting report on **Smart Search Over Enciphered Data In Cloud Computing**. Completing a task is never a one-man effort. It is often a result of invaluable contribution of a number of individuals in direct or indirect manner. I would like to express deepest appreciation towards **Dr. V.V.Dixit**, Principal RMDSSOE, Warje, Pune 58 and **Prof. V. M.Lomte**, HOD Computer Department, whose invaluable guidance supported me in completing this report.

I am profoundly grateful to **Ms.Nikhita Nerkar** for her expert guidance and continuous encouragement throughout to see that this project report rights its target since its commencement to its completion.

At last I want to express my sincere heartfelt gratitude to all the staff members of Computer Engineering Department who helped me directly or indirectly during this course of work.

**Gauri Kavitkar**

**Sarvesh Kher**

**Gaurav Mahendrakar**

**Mayuresh Patil**

Date: / / 2019

# ABSTRACT

With the advancement of information technologies particularly cloud storage used outsourcing data. Now a days users store a large amount of data on the cloud but its untrusted and we store secure data on the cloud.

The concept of searchable encryption provides a promising direction in solving the privacy problem when outsourcing data to the cloud. Such schemes allow users to store their data in encryption from at an untrusted server, and then delegate the server to search on their behalf by issuing a private key and encrypted search index. This proposed system is more efficient than previous systems in security as well as response time.

**Keywords-** ***Searchable Encryption, Data Outsourcing, Cloud Computing, TF-IDF algorithm, Search Index.***

List of Abbreviation

1. ECC - Elliptic-curve cryptography

2. CA - Certificate Authority

3. AES - Advance Encryption Standard

4. TF - Terms Frequency

5. IDF - Inverse Document Frequency

# List of Figures

# Contents

# Chapter 1

# INTRODUCTION

## 1.1 OVERVIEW

With the development of cloud computing, more and more information data can be stored to the public cloud to take the advantage of economical savings and right to use. On the other hand, the private information has to be encrypted to give assurance about the security. For instance, a user Alice may outsource her data at Dropbox and distribute them with her friends, in the mean time she may also have access to her friends data. Because of the private nature of personal data, there is an intrinsic requirement for a user to selectively share her data with different recipients. In practice, what a user can do is to set some access control policies and then rely on the cloud server (e.g. Dropbox) to enforce them. unluckily, this approach is not sensible because of two reasons. First one is that the users have no means to stop the server from accessing their information. The other is that, even if the server is begin, it may also be forced to share users data with other parties.

To preserve privacy, the datasets are usually encrypted before outsourcing. To put into practice proficient data exploitation, over encrypted cloud data has been a great challenge. The existing solutions depended entirely on the submitted query keyword and do not consider the semantics of keyword. Thus the search schemes are not intelligent and also omit some semantically related documents. We also store large amount of data on the cloud which cannot be trusted and you may have loss of data or any other person can have access to your data due the weak security on cloud which may cause problem.

To overcome this problem the concept searchable Encryption provides security and secures data when outsourcing to cloud. Such schemes allow users to store their data in encrypted form at an untrusted server, and then delegate the server to search on their behalf by issuing a trapdoor (i.e. encrypted keyword). A detailed survey of searchable encryption schemes can be found in [8][9]. As to the specic setting where multiple users store and share their data with each other in the cloud, we need a new primitive, namely multi-party searchable encryption (MPSE)[10] schemes in the symmetric setting. MPSE can be regarded as a multi-party version of the symmetric searchable encryption proposed by Qiang Tang et al. in [10] a MPSE scheme allows every user to build an encrypted index for each of her documents and store it at a cloud server. The index contains a list of encrypted keywords, as well as some authorization information which selectively authorizes other users to search over this index.

In the proposed system the user need not to share the private key with the server as a result server cannot decrypt the data of the user and the information security remains high. User can generate his private key after accessing the public key as a result can have access each different document with different key therefore the privacy for data is enhanced. Authorization information along with index file of each document helps to detect malicious users and prevent them from accessing users private data. Only person sharing and receiving the data will be aware of the information shared and no third party is involved.

The TD IF algorithm is used that helps in time efficiency. Server does not play major role in this system as it is used only to maintain clusters that will occur during the frequent search. The proposed system is more secured and has no time complexity issues than other systems.

## 1.2 MOTIVATION

Today cloud can be used to store large amount of data which cannot be trusted and there is chances of loss of data or any other person can have access to your data due the weak security on cloud which may cause problem.The existing solutions depended entirely on the submitted query keyword and do not consider the semantics of keyword. In the proposed system the user need not to share the private key with the server as a result server cannot decrypt the data of the user and the information security remains high.

## 1.3 PROBLEM DEFINITION AND OBJECTIVES

To implement a system i.e.“Semantic-aware Searching over Encrypted Data for Cloud Computing” which can provide higher level of security in searching valuable information which is to be shared by multiple users.

**Objectives**

- To provide security to the data shared over the cloud.
- To reduce the time complexity for response of a system by using keyword search.
- To provide authenticated access to the data over cloud by sharing the keys.
- To build a secure network and keep the privacy of the data.

## 1.4 PROJECT SCOPE & LIMITATIONS

In this system firstly documents are stored on cloud along with mapping index in the encrypted format. Key’s ae generated for single users document. a single key having multiple authorization codes will be used for decryption. After sign in by the user role and department of the user will be extracted from login information which is used for extracting users trapdoor key from the key. That trapdoor key is used for searching and locating the user document. For encrypting and decrypting the document AES algorithm is used. For searching the document TF-IDF algorithm technique is used which will provide excellent time efficiency.

**Limitation**

- User need to carry private when ever wants to access data

## 1.5 METHODOLOGIES OF PROBLEM SOLVING

- This platform is rapidly growing with users need which overcomes the issues of security which lead to the poor efficiency. Software project estimation is form of problem solving.

- The complex software is hard to estimate hence it is divided into smaller pieces. The estimation of project will be correct only when the estimation of size of the project is correct. In the context of project planning size refers to quant able outcome of project.

- Here, the direct approach is selected and hence, the size is estimated in Line of Codes. The feasibility study comprise of an initial investigation into personnel will be required. Feasibility study will enable us to make informed and straightforward choice at crucial points while developing phase. All projects are feasible given unlimited times and resources. But, the development of computer based system is more likely to be plagued to scarcity of resources .It is both essential and prudent to evaluate the feasibility of project at earliest possible time.

- We are developing an efficient method that provides security and secures data when outsourcing to cloud.

# Chapter 2

# LITERATURE SURVEY

Curtmola et al.[1]proposed the concept of multi-user searchable encryptionschemes, where a user can authorize multiple other users tosearch her encrypted data. However, the proposed primitivedoes not take into account the fact that the same user may alsobe authorized to search other users data and the correspondingsecurity issues. As a result, the primitive from [1] offers asolution for a much more simplified problem than ours, andit seems not trivial to construct a scalable solution for ourproblem based on their scheme.

Dividing task into the three entities, Group Manager, Opening Manager and Revocation Manager, [3] toincrease the privacy of the user by dividing the work of the group manager into three entities. Group Manager can only create group and add members in group but does not possesses power to open any signature. The Open Manager possesses a special key which can be used only to open a signed message.

The main idea behind our scheme is that the secret key of the group [8] is split into two parts by GM, one part is given to the user as his group membership secret key, and the other is given to SEM. Neither the group member nor SEM can sign a message without the others help. To revoke the membership of a group member, GM need only ask SEM not to provide the group member partial signatures any more.

The paper [5] propose a systematic solution, which refers to as QDMiner, to automatically mine query facets by extracting and grouping frequent lists from free text, HTML tags, and repeat regions within top search results. Experimental results show that a large number of lists do exist and useful query facets can be mined by QDMiner. They further analyze the problem of list duplication and nd better query facets can be mined by modeling ne-grained similarities between lists and penalizing the duplicated lists.

In [10] paper, the author considers objects that are tagged with keywords and are embedded in a vector space. For these datasets, they study queries that ask for the tightest groups of points satisfying a given set of keywords. It proposes a novel method called ProMiSH (Projection and Multi-Scale Hashing) that uses random projection and hash-based index structures and achieves high scalability and speedup. Also, they present an exact and an approximate version of the algorithm. The experimental results on real and synthetic datasets show that ProMiSH has up to 60 times of speedup over state-of-the-art tree-based techniques.

Long et al. [2] proposed algorithms toretrieve a group of spatial web objects such that the groupskeywords cover the querys keywords and the objects in thegroup are nearest to the query location and have the lowestinter-object distances. Other related queries include aggregatenearest keyword search in spatial databases. First,existing works mainly focus on the type of queries wherethe coordinates of query points are known [5], [2].

Subhra Mishra and TilakRajanSahoo[7] The scheme implemented by us provides these features. The use of elliptic curve cryptography increases the security the scheme by providing desired security level that is achieved by significantly smaller keys in elliptic curve system than in its counterpart- RSA system. Another significant advantage being in general, the algorithms used for encryption and decryption in ECC schemes are faster and can be run on machines that are less efficient.

The secret key of the group is split into two parts by GM, one part is given to the user as his group membership secret key, and the other is given to SEM. Neither the group member nor SEM can sign a message without the others help. To revoke the membership of a group member, GM need only ask SEM not to provide the group member partial signatures any more.

# Chapter 3

# SOFTWARE REQUIREMENTS SPECIFICATION

## 3.1 ASSUMPTIONS AND DEPENDENCIES

Let us Assume:

- The user must have a basic knowledge of computer and handling web application.
- Firewall for the system from where the data is shared.
- It is assumed that the maintenance of the database will be assigned to the authorized person only.

**Dependencies:**

- Requires internet connectivity
- Only Administrators will be able to edit main configurations.
- Admin need to register the user first.

## 3.2 SYSTEM FEATURE 1 (FUNCTIONAL REQUIREMENTS)

- Only authenticated client can have access to the system.

## 3.3 SYSTEM FEATURE 2 (FUNCTIONAL REQUIREMENTS)

**User Module**

- User sign in to the system
- User shares the data secured using private key through the system.
- User request for searching the document.

**Cloud**

- Store the data with index.
- Store User credentials.
- Store Shared data.

**System**

- Data and index encryption can be done using AES algorithm.
- Document searching is performed using encrypted index.
- Content based filtering with TF-IDF is used to provide quick and easy search for the document.
- Key management is done with ECC algorithm.

## 3.4 EXTERNAL INTERFACE REQUIREMENTS

### 3.4.1 User Interfaces

- User interface will be provided with the required information.
- User interface will provide good look and feel effect so that it will user friendly
- And he or she can operate system very efficiently.
- Various Tools will be available on the user interface which the user can operate.

### 3.4.2 Hardware Interfaces

Server side System will be windows based supporting versions windows 07 onwards. The minimum configuration required on server platform.

- **System :** Intel I3 or above
- **RAM :** 4GB RAM 80GB memory

### 3.4.3 Software Interfaces

- **IDE :** Eclipse
- **Platform :** Microsoft Windows 7 Professional or greater
- **Language :** Java
- **Database :** MySQL

### 3.4.4 Communication Interfaces

- Communication Interface process is intended to give an approach to archive and track extend interfaces from Planning stage (FEP) to the end of the project.
- The system use the HTTP protocol for communication over the internet and for the intranet communication will be through TCP/IP protocol suite.

## 3.5 NONFUNCTIONAL REQUIREMENTS

### 3.5.1 Performance Requirements

**High Speed**

The system should process the requested task in parallel for various activities to give a quick response then the system must wait for process completion.

**Accuracy**

The system should correctly execute the process; i.e. display the result i.e according to the particular parameter.
System output should be in user required format.

**Interoperability**

System should have the ability to exchange information and communicate with internal and external applications and systems. It must be able exchange information both internally and externally.

**Response Time:**

The response time of the system should be deterministic at all times and very low, i.e it should meet every deadline. Thus, the system will work in real time.

### 3.5.2 Safety Requirements

- The data safety must be ensured by arranging for a secure and reliable transmission media. The source and destination information must be entered correctly to avoid any misuse or malfunctioning.
- The source and destination information must be entered correctly to avoid any misuse or malfunctioning.
- Safety requirements against the natural disaster and accidents.
- Failures due to technical issues.

### 3.5.3 Security Requirements

- All the user details shall be accessible to only high authority persons.
- Access will be controlled with usernames and passwords.

### 3.5.4 Software Quality Attributes

- Maintainable software should have

- Encourage in-code documentation (XML docs in javadoc, etc.)
- use a wiki to maintain the documentation
- Unit Tests = Good for documenting specifications
- Comments = Good for documenting design decisions.
- Unit Tests + Comments = Good for documenting specifications and design decisions. = Easily maintainable software.
- Faster feedback from any changes made to the system
- Providing better transparency into the changes happening to the system
- Propagating environmental changes and code changes more rapidly while maintaining control

## 3.6 SYSTEM REQUIREMENTS

### 3.6.1 Database Requirements

The database is required to be created and maintained in MySQL Server. Stored procedures are also created to retrieve and operate on data.

### 3.6.2 Hardware Requirements

The minimum configuration required on server platform.

- **System :** Intel I4 or above
- **RAM :** 256MB RAM 80 GB memories

### 3.6.3 Software Requirements

**Eclipse**

- Eclipse is an open source community whose projects building tools and frameworks are used for creating general purpose application. The most popular usage of Eclipse is as a Java development environment.

- Eclipse is an open source community, whose projects are focused on building an open development platform comprised of extensible frameworks, tools, and runtimes for building, deploying and managing software across the lifecycle. The Eclipse Foundation is a not-for-profit, member supported a corporation that hosts the Eclipse projects and helps cultivate both an open source community and an ecosystem of complementary products and service
- The independent not-for-profit corporation was created to allow a vendor-neutral and open, transparent community to be established around Eclipse. Today, the Eclipse community consists of individuals and organizations from a cross-section of the software industry.
- In general, the Eclipse Foundation provides four services to the Eclipse community:
  - IT Infrastructure
  - IP Management
  - Development Process, and
  - Ecosystem Development.

  Full-time staff is associated with each of these areas<br>
  and work with the greater Eclipse community to assist in meeting the needs of the stakeholders.

**Java**

- The Java Development Kit (JDK) is a software development environment used for developing Java applications and applets. It includes the Java Runtime Environment (JRE), an interpreter/loader (java), a compiler(javac), an archiver (jar), a documentation generator (javadoc)and other tools needed in Java development.
- A Java virtual machine(JVM) is an abstract computing machine that enables a computer to run a Java program. There are three notions of the JVM: specification, implementation, and instance. The specification is a document that formally describes what is required of a JVM implementation. Having a single specification ensures all implementations are interoperable.

- A JVM implementation is a computer program that meets the requirements of the JVM specification. An instance of a JVM is an implementation running in a process that executes a computer program compiled into Java bytecode.

**MySQL**

- Java web application will require storing large amounts of metadata and keep data organized. Therefore there was a need to host a Java web application with MySQL. A few other benefits of using MySQL as opposed to other database software for your Java hosting include:

- State-of-the-art security: MySQLs reputation as the safest relational database currently in use makes it ideal for e-commerce sites that handle frequent online transactions and other sensitive data.

- High-quality performance: Built to handle the most demanding websites with the heaviest traffic, its not bogged down by high usage. Even when its used by traffic-heavy sites like Twitter and Facebook, MySQL maintains its lightning fast performance speeds.

- More uptime: MySQL guarantees 100% uptime so that you never have to worry about surprise software crashes.

- Easy maintenance: Because its open-source, the software is constantly being upgraded and debugged, which means less maintenance for you to worry about all you have to worry about your Java site or web application.

- Its used everywhere: MySQLs popularity actually doubles as a benefit because its an industry standard, its compatible with almost any operating system you can think of. Following are the basic steps that are needed to follow for setting up dedicated hosting Server:
  1. Build your own dedicated server
  2. Install Apache Tomcat
  3. Install the latest version of MySQL (versions are available for Windows, Linux, and Mac)

4. Configure and test your MySQL installation

### Apache Tomcat Server

- What is not in doubt though is that it is currently one of the most widely used application servers in the market. As a matter of fact, many of todays applications and virtually all web services can be built on top of Tomcat with a variety of add-ons and pluggable services readily available in the market. It is no secret that many developers acknowledge that Tomcat is usually a much better choice to build todays deployment and development architectures than other servers.
- To put it simply, Tomcat provides the environment in which Java servlets are executed and web page client requests are processed.
- Another main advantage of the product is the ease of installing and configuring the application. Typically, this can be done in less than twenty (20) minutes. It is also worth mentioning that deploying web applications to Tomcat is also very easy and simple. Apache Tomcat is an open source web server that is developed by the Apache software foundation. It is designed to run all Java web applications completely produced and taken care by Apache System. It offers HTTP protocol through users from anywhere can connect with the server by its URL and access the Java application which is deployed in it. There is a built-in web container called Catalina in the tomcat bin directory. It loads all HTTP related request and has the privilege to instantiate the GET and POST method's object.
- It basically makes our Java Web applications to run on host and server-based system and it is configured on localhost port 8080. It generally runs JSP, Servlet etc. Hosting Tomcat on the dedicated server isnt better just because it offers superior customization and control some of the other advantages of private Tomcat are:
- Availability: When you have your own instance of Tomcat, you dont have to worry about other applications hogging the servlet container and slowing it down. The only Java application that will be running is your own.

- Manager Access: You have full access to administrative and managerial functions, giving you full control over individual applications.

- Easy Deployment: Using the management tools private Tomcat hosting provides, you can deploy WAR and JAR files quickly and efficiently through Tomcat Manager.

- Flexibility: Private Tomcat hosting gives you the freedom to choose whichever version of Tomcat you want to host with so that you can guarantee the best possible hosting environment for your Java application.

## 3.7 ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

### Iterative SDLC Model

The Iterative SDLC model does not need the full list of requirements before the project starts. The development process may start with the requirements to the functional part, which can be expanded later. The process is repetitive, allowing to make new versions of the product for every cycle. Every iteration includes the development of a separate component of the system, and after that, this component is added to the functional developed earlier.

Speaking with math terminology, the iterative model is a realization of the sequential approximation method; that means a gradual closeness to the planned final product shape.The key to a successful use of an iterative software development life cycle is rigorous validation of requirements, and verification and testing of each version of the software against those requirements within each cycle of the model. As the software evolves through successive cycles, tests must be repeated and extended to verify each version of the software.
The major steps of the SDLC model are given below:

- Requirement gathering: All the functional and non-functional requirements of the project were identified. Interaction with the users and all other stakeholders of the project was conducted to identify all the requirements starting from important

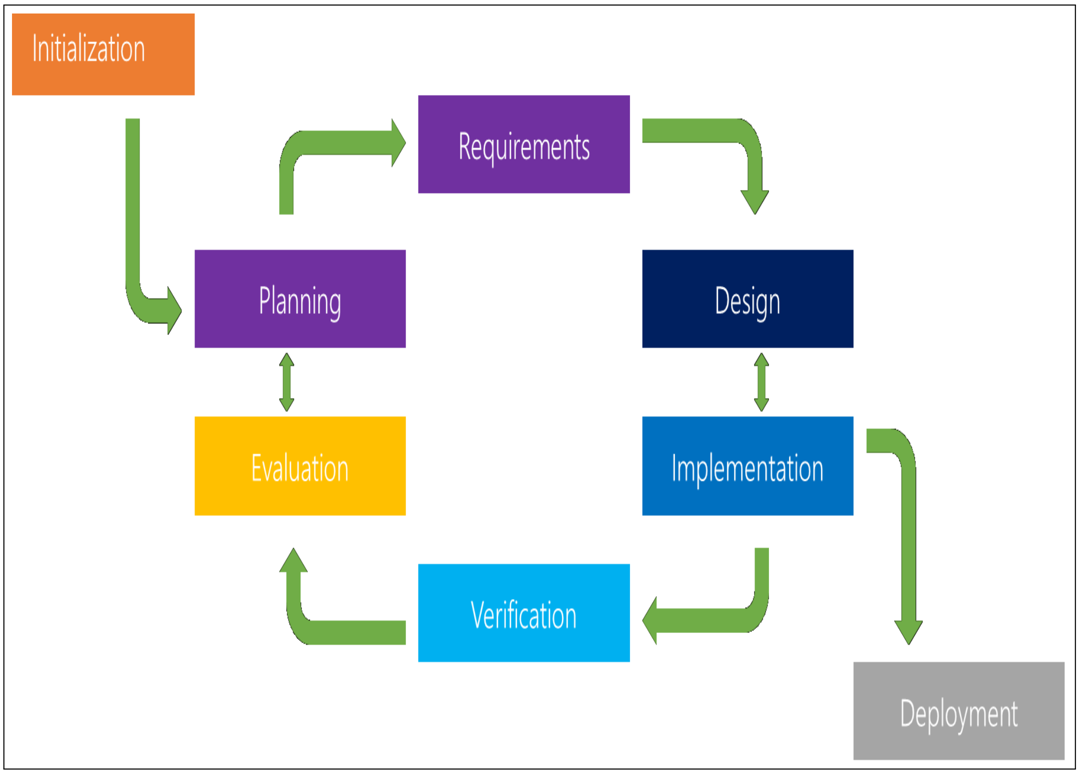


Figure 3.1: Iterative SDLC Model

features like maintaining audit trail, security parameters etc. to the very basic features like the look and the feel of user interface. The different requirements mainly fall into categories:

1. System features
2. Security parameters
3. User requirements
4. Administrator requirements
5. User interface

- Design: The first step was database design. A complete database required for the implementation of this project was designed. The second step was project design. The project was designed based on a framework. The framework uses three layers:

  a. Business entities layer: It identifies all the entities used in the project.

  b. Business logic layer: This layer operates on the business entity to achieve the goals.

c. Data access layer: This layer serves as an interface between backend and the services.

- Construction: All modules and user interface was built in this step. Development was done using Java. Database was constructed in MySQL.
- Integration and system testing: All the modules were integrated together. The user interface was integrated with the modules which made the use web services. Data flow originated from the database built in MySQL. In testing phase project was tested and debugged. Various test cases were developed and the project was tested at the developers end as well as users end. Debugging was done to discover errors and exception which were corrected.
- Installation and maintenance: Our system is installed on one dedicated machine and it is accessible to admin and all authenticated users. Maintenance of our system is done on regular basis. New requirements and features can be added as and when required as long as they do not conflict with the existing features

# Chapter 4

# SYSTEM DESIGN

## 4.1 SYSTEM ARCHITECTURE

The proposed system Users do not need to share private key with server, so server cannot decrypt users data and data confidentiality remains high. User can self-generate his private key after choosing a public key. User can encrypt each different document using different public keys by this way user can provide more security to each of his text document. So the new proposed scheme provides higher level of security in searching valuable information which is to be shared by multiple users. Authorization information along with index file of each document helps to detect malicious users and prevent them from accessing users private data. Figure 4.1 shows the architecture of the proposed system.

- **Store documents on cloud :**
  Store documents on cloud with mapping the index. All Documents are encrypted along with the search index. Search index is given to each document for fast searching.

- **Key Exchange with User:**
  Generate keys for single users document, a single key having multiple authorization codes will be used for decryption purpose. Each user will have to sign in to our system; its role and department will be extracted from login information. Department and role will be used for extracting users trapdoor key from the key. Once the trapdoor key is extracted, it will used for searching and locating the

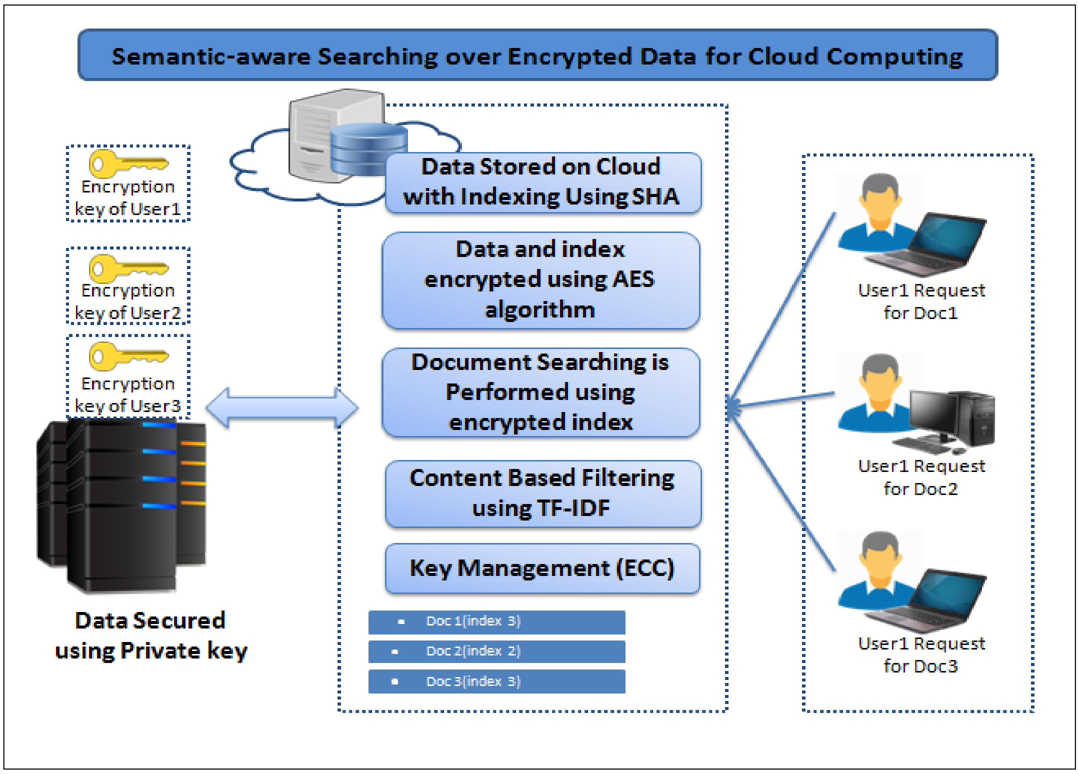

Semantic-aware Searching over Encrypted Data for Cloud Computing
Encryption key of User1
Encryption key of User2
Encryption key of User3
Data Secured using Private key
Data Stored on Cloud with Indexing Using SHA
Data and index encrypted using AES algorithm
Document Searching is Performed using encrypted index
Content Based Filtering using TF-IDF
Key Management (ECC)
Doc 1(index 3)
Doc 2(index 2)
Doc 3(index 3)
User1 Request for Doc1
User1 Request for Doc2
User1 Request for Doc3

Figure 4.1: System Architecture

user document.

- **Key Management (ECC Algorithm);**

  Key management done with the help of ECC Algorithm. Apply ECC Algorithm for public private key. Data will encrypt using public private key.

- **Multi-party searchable encryption:**

  When user gets request for searching document. That time he/she sends search index according to the cloud server. Once the data will searched then data will decrypt using AES algorithm.

- **Content Based Filtering using TF-IDF:**

  - Searching data is divided into several attributes example Item U may be having attributes A1,A2,A3,A4...An
  - We have several items in the database may be U1, U2, U3...UN.
  - Each items attributes are compared to rest of the items in the database and a cumulative score is calculated based on their similarities.

Hence a algorithm is used to match these attributes example if U1 has A1 as A,B,C and U2 has A1 as B,C then their matching score would be U1(A1)intersection U2(A1)/ # of U1(A1).

## 4.2 MATHEMATICAL MODEL

$$y^2 = x^3 + ax + b$$

For current cryptographic purposes, an elliptic curve is a plane curve over a finite field which consists of the points satisfying the equation $y^2 = x^3 + ax + b$ along with a distinguished point at infinity denoted . This set together with the group operation of ellitic curves is an abelian group with the point at infinity as identity element. The structure of the group is inherited from the divisor group of the underlying algebric variety.

## 4.3 DATA FLOW DIAGRAMS

A data flow diagram (DFD) is a graphical representation of the flow of data through an information system, modeling its process aspects. It shows data is processed by a system in terms of inputs and outputs.

### 4.3.1 DFD Level-0

It only contains one process node (Process 0) that generalizes the function of the entire system in relationship to external entities.

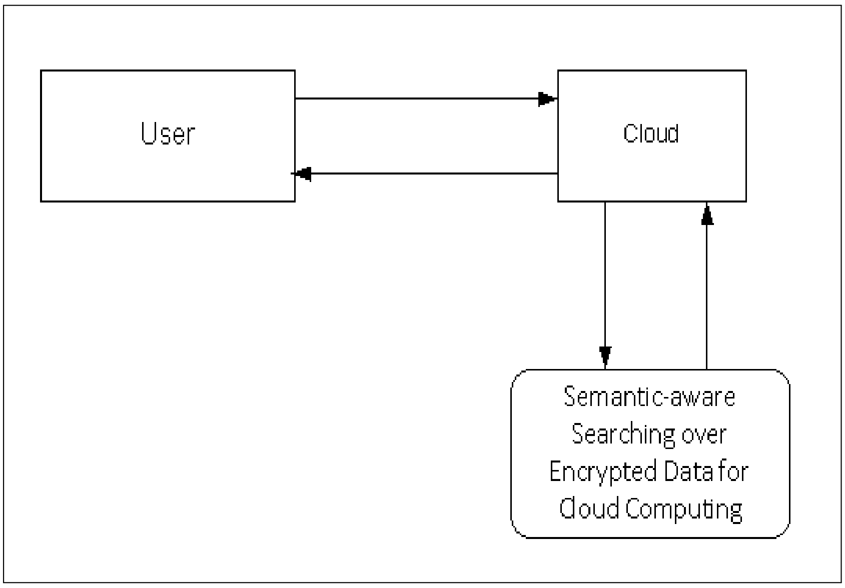

User
Cloud
Semantic-aware Searching over Encrypted Data for Cloud Computing

Figure 4.2: DFD Level-0

### 4.3.2 DFD Level-1

DFD level 1 diagram expands the DFD 0 and shows the detailed flow of the proposed system.

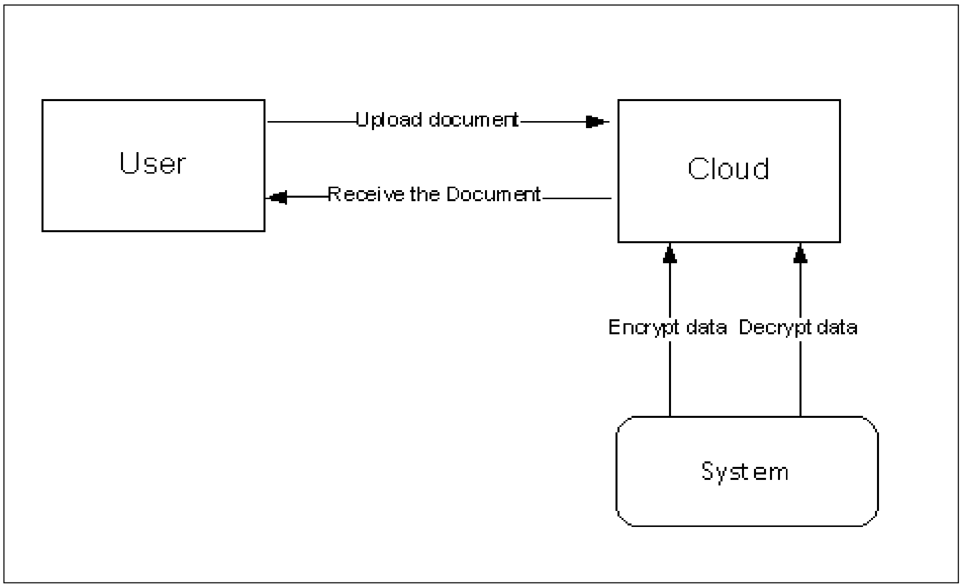


Figure 4.3: DFD Level-1

### 4.3.3 DFD Level-2

DFD level 2 diagram expands the DFD 1 and shows the detailed flow in the proposed system. It shows the different processes that take place to perform the authentication.

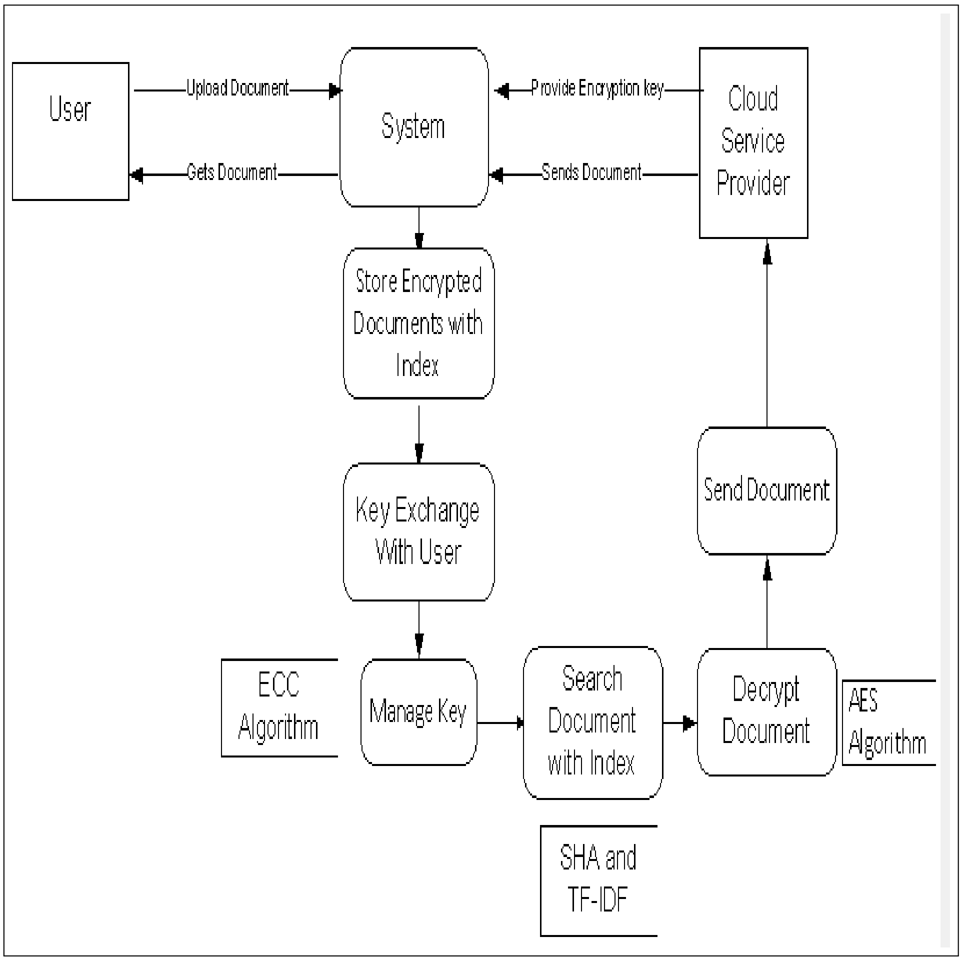


Figure 4.4: DFD Level-2

## 4.4 ENTITY RELATIONSHIP DIAGRAMS

Data objects and their major attributes and relationships among data objects are described using an ER - like form.ER diagram is a data model for describing the data or information aspects of a software system. The main components of ER models are entities and the relationships that exist among them. The various entities are system and admin.

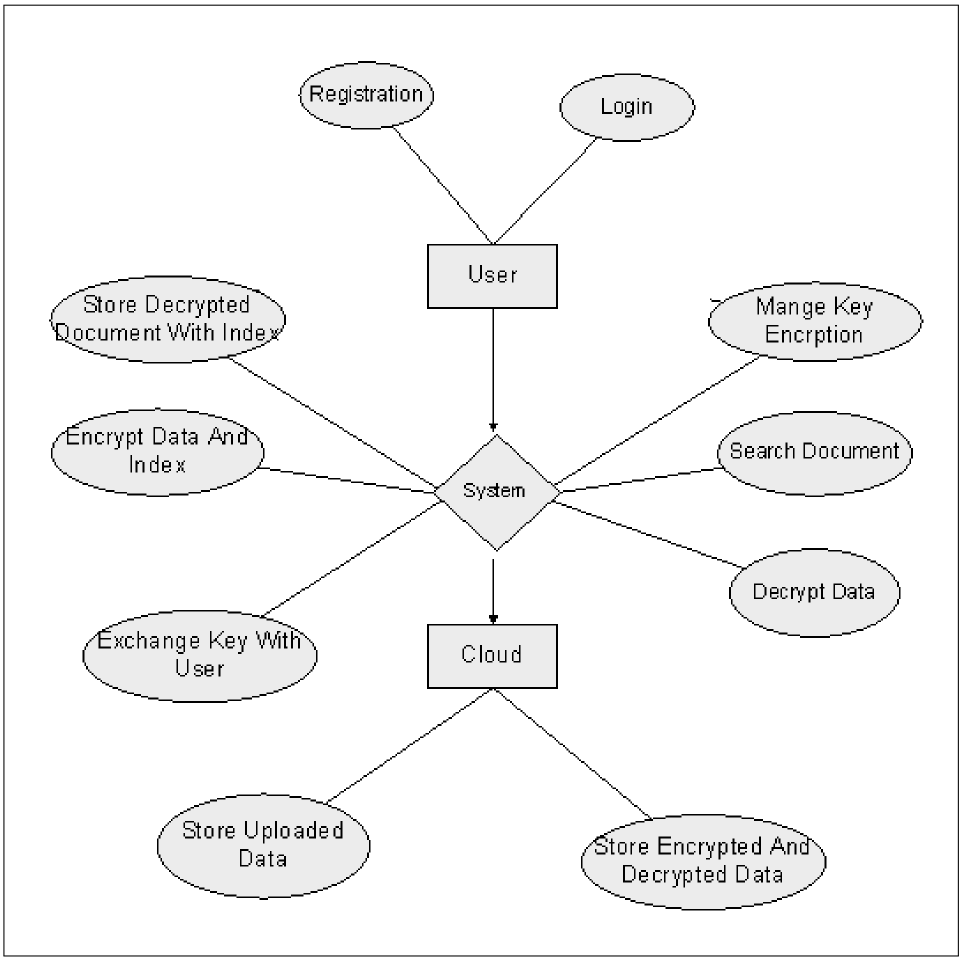


Figure 4.5: ER Diagram

## 4.5 UML DIAGRAMS

### 4.5.1 Use Case Diagram

Use case diagram is a simple representation of a users interaction with the system that shows the relationship between the user and the different use cases in which the user is involved.Here the actors are manufacturer, distributer, customer and system.

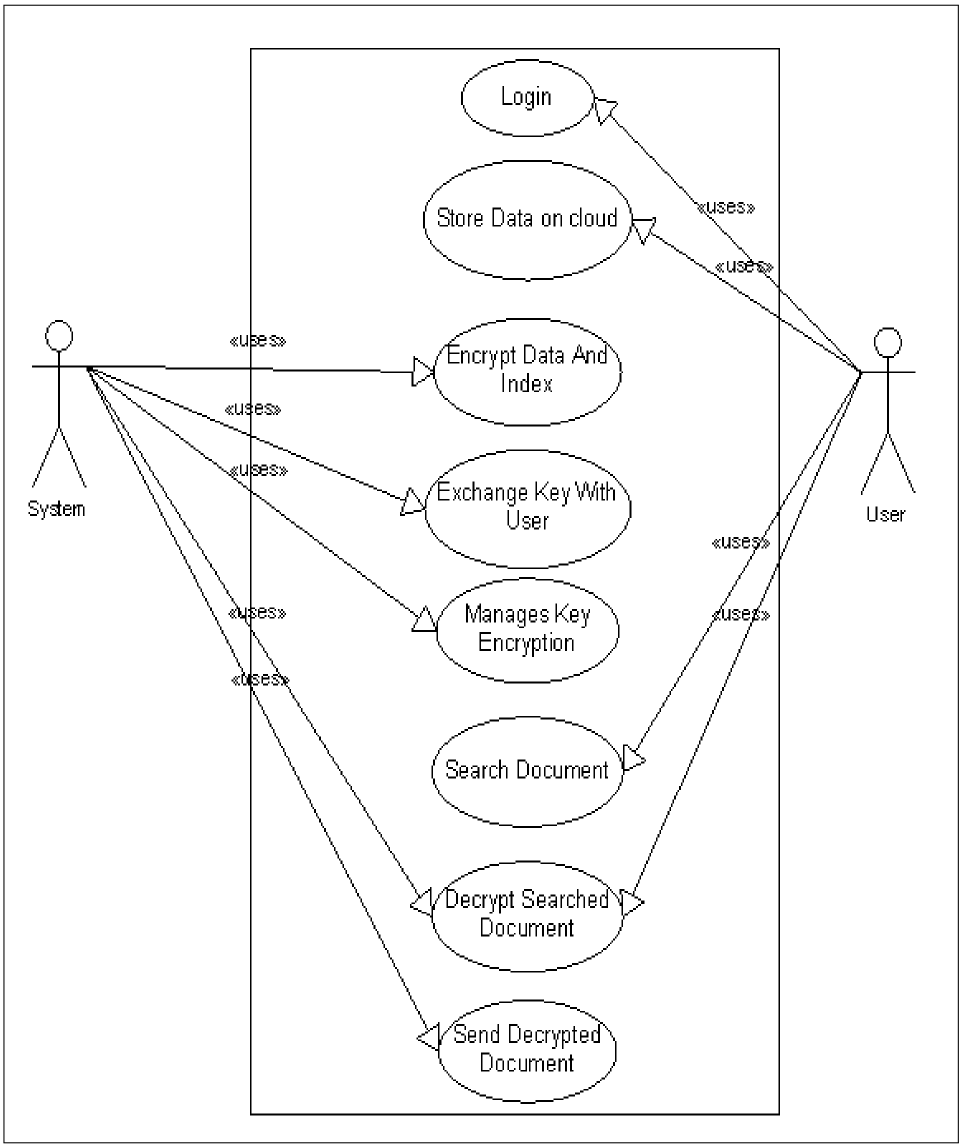


Figure 4.6: Use Case Diagram

### 4.5.2 Sequence Diagram

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. Sequence diagrams are sometimes called event diagrams or event scenarios. The sequence diagram for the proposed system shows the interaction in between admin, system and database.

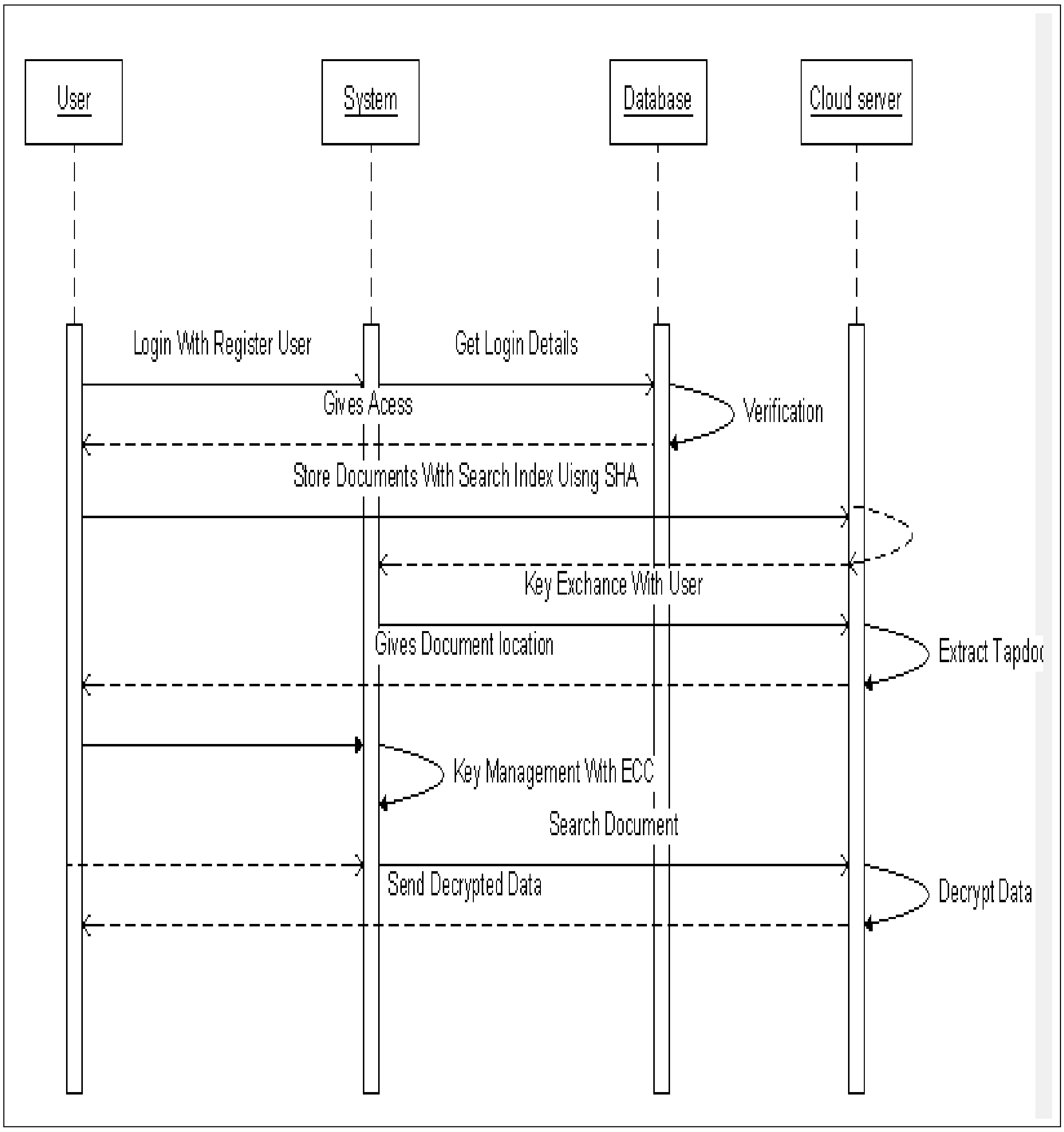


Figure 4.7: Sequence Diagram

### 4.5.3 Class Diagram

Class diagram is a type of structure diagram that shows the structure of the classes, attributes, operations and relationship among them. Given below is the class diagram of the proposed system which shows 8 classes such as user, system, database.

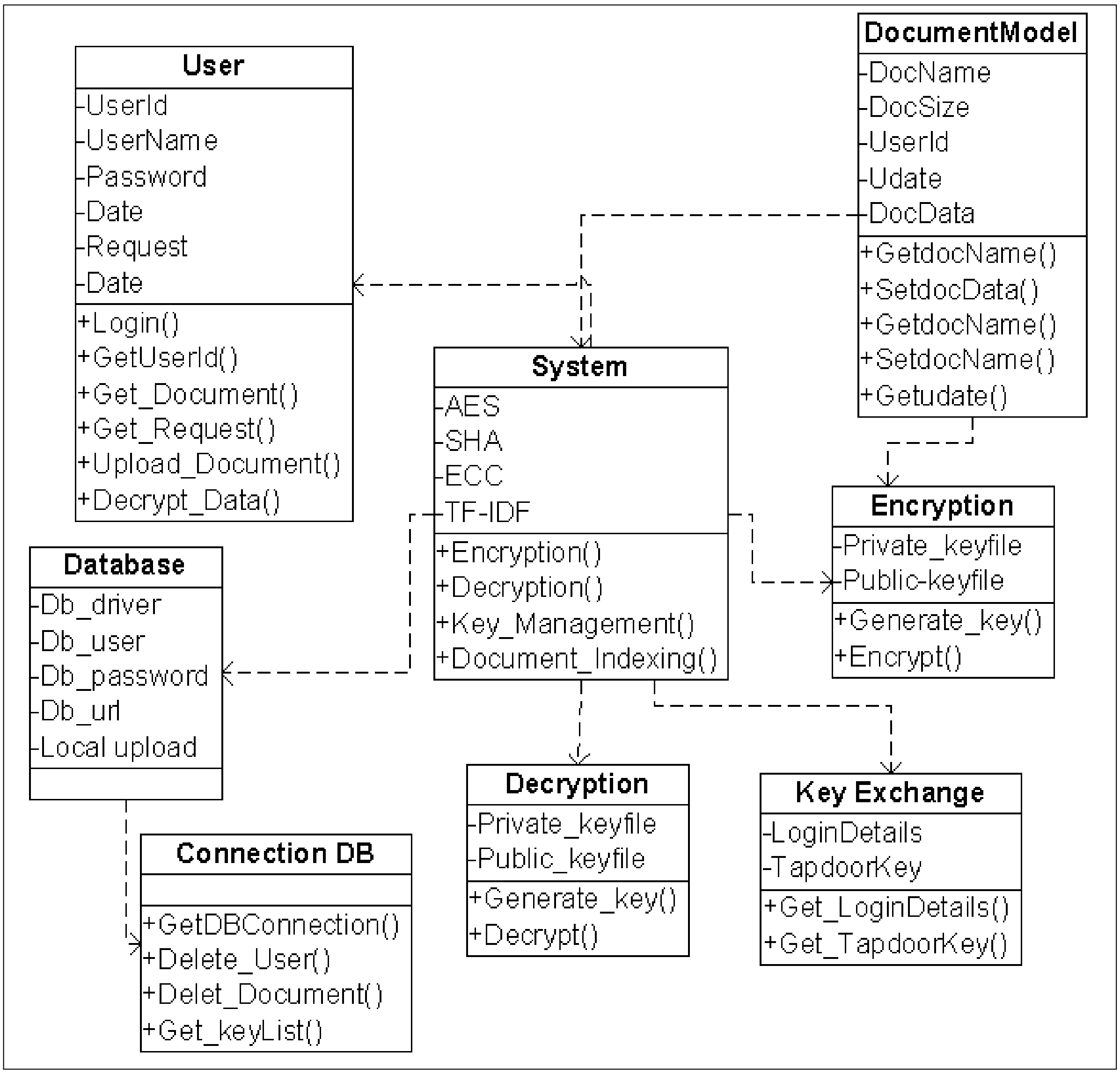


Figure 4.8: Class Diagram

### 4.5.4 Component Diagram

A component diagram depicts how components are wired together to form larger components and or software systems. A component is something required to execute a stereotype function. Examples of stereotypes in components include executable, documents, database, tables, files.

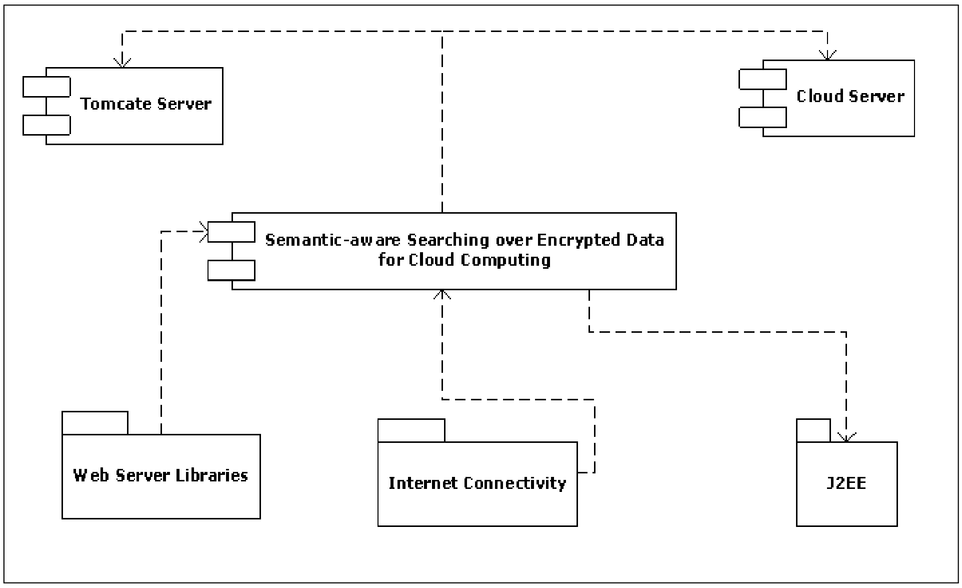


Figure 4.9: Component Diagram

### 4.5.5 Deployment Diagram

Deployment diagrams are used to visualize the topology of the physical components of a system where the software components are deployed. The deployment diagram for the proposed system shows below. It shows the physical or the hardware components on which the software components. The physical components include the Server, Client, Windows JVM and the Database.

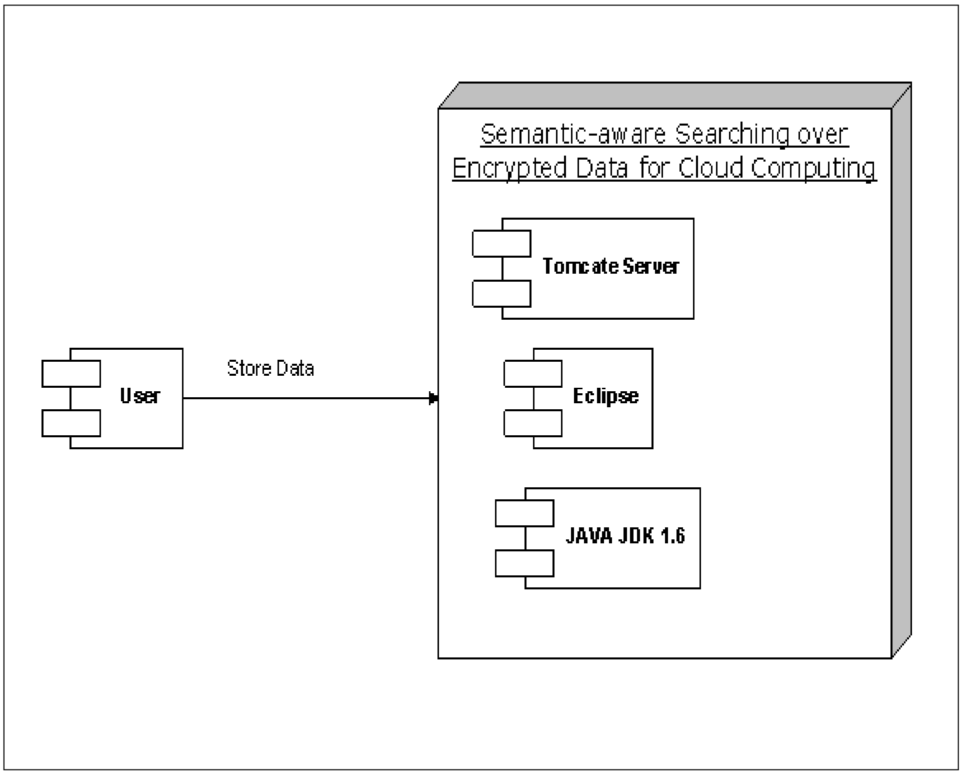

Semantic-aware Searching over Encrypted Data for Cloud Computing
Tomcate Server
User
Store Data
Eclipse
JAVA JDK 1.6

Figure 4.10: Deployment Diagram

# Chapter 5

# PROJECT PLAN

## 5.1 PROJECT ESTIMATE

Use of Waterfall model and associated streams derived for estimation.

### 5.1.1 Reconciled Estimates

**Cost Estimate**

The initial cost estimate of the project before beginning the implementation process is INR 15000 for in-house resources. This cost may vary. This estimate is subject to change according to the availability and/or need of a particular item.

**Time Estimates**

The initial time estimate for the complete implementation of the primary objectives is 45-50 days depending on the schedule of the developers. The secondary objectives require an additional of 25 days to be completed. Also, depending on the stage of development, the testing and debugging would require an additional of 15 days.

### 5.1.2 Project Resources

**Hardware**

1. System: Intel P4 or above
2. RAM: 256MB RAM 80 GB memories.

**Software**

- **Platform:** Eclipse
- **Database :** MYSQL
- **Platform :** Microsoft Windows 7 Professional or greater
- **Language :** Java

## 5.2 RISK MANAGEMENT

During different phases of project there can be several threats like uncertainty in financial markets, threats from project failures. These threats if not at- tended can later cause poor performance and project failure. By knowing predictable risks, the project manager takes first step towards avoiding them and controlling them when necessary. Two types of different risks are:
**Generic risks:** Generic risks are a potential threat to every software project.
**Product-specific risks:** Product-specific risks can be identified only by those with a clear understanding of the technology, the people, and the environment that is specific to the project at hand.

### 5.2.1 Risk Identification

It is the process of determining potential threats which can later harm the performance of the project. One method of identifying risks is to create a risk item checklist. The checklist can be used for risk identification and focuses on some subset of known and predictable risks in the following generic subcategories:

- Product size : Risks associated with the overall size of the software to be built or modified.
- Business Impact: Risks associated with constraints imposed by management or the market place.
- Customer characteristics: Risks associated with the sophistication of the customer and the developers ability to communicate with customer in a timely manner.

- Process definition: Risks associated with the degree to which the soft- ware process has been defined and is followed by the development organization.
- Development Environment: Risks associated with the ability and quality of the tools to be used to build the product.
- Technology to be built: Risks associated with the complexity of the system to be built.
- Staff size and experience: Risks associated with the overall technical and project experience of the software engineers who will do the work.

### 5.2.2 Risk Analysis

Risk management is concerned with identifying risks and drawing up plans to minimize their effect on a project. A risk must also have a probability. It must be a chance to happen or it is not a risk. The risks for project can be analyzed within the constraint of time and quality.

| ID | Risk Description | Probability | Impact | | |
|---|---|---|---|---|---|
| | | | Schedule | Quality | Overall |
| 1 | Correctness | Low | Low | High | Low |
| 2 | Availability | High | Low | High | High |

Table 5.1: Risk Table

### 5.2.3 Overview Of Risk Mitigation, Monitoring, Management

Following are the details for each risk.

## 5.3 PROJECT SCHEDULE

### 5.3.1 Project Task Set

Major Tasks in the Project stages are:

| | |
|---|---|
| Risk ID | 1 |
| Risk Description | Third party access |
| Category | Networking Environment |
| Source | Internet. |
| Probability | High |
| Impact | High |
| Response | Mitigate |
| Strategy | Break security |
| Risk Status | Occurred |

| | |
|---|---|
| Risk ID | 2 |
| Risk Description | User can make fake profle |
| Category | Requirements |
| Source | Software Design Specification documentation review. |
| Probability | Low |
| Impact | High |
| Response | Mitigate |
| Strategy | Better testing will resolve this issue. |
| Risk Status | Identified |

| | |
|---|---|
| Risk ID | 3 |
| Risk Description | Server crash |
| Category | Technology |
| Source | This was identified during early development and testing. |
| Probability | Low |
| Impact | Very High |
| Response | Accept |
| Strategy | Example Running Service Registry behind proxy balancer |
| Risk Status | Identified |

- Task 2 : Literature Survey
- Task 3 : Applications and Objectives
- Task 4 : Platform/Technology Selection
- Task 5 : Internal Presentation - 1
- Task 6 : Study Of Algorithms
- Task 7 : Mathematical Model
- Task 8 : Software Requirements Specification
- Task 9 : UML Diagrams
- Task 10 : Problem Definition using NP Hard/ NP Complete
- Task 11 : System Architecture
- Task 12 : Testing phase
- Task 13 : Internal Presentation - 2
- Task 14 : Report Preparation
- Task 15 : Installation
- Task 16 : Overview of Project Model
- Task 17 : Construction of GUI
- Task 18 : Module Identification
- Task 19 : Module 1 - User Authentication
- Task 20 : Module 2 - Database generation
- Task 21 : Module 3 - Connection of GUI to Database
- Task 22 : Module 4 - Testing and Result
- Task 23 : Test Planning
- Task 24 : Testing

- Task 25 : Poster Presentation
- Task 26 : Research Of Journals For Final Report

### 5.3.2 Task Network

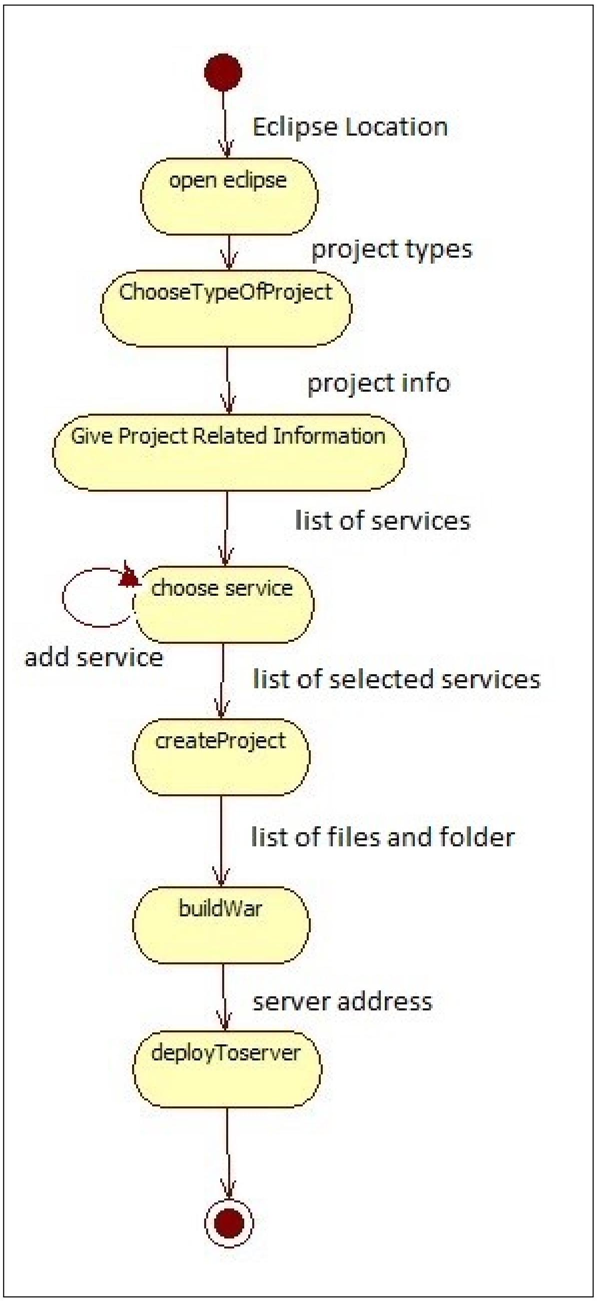
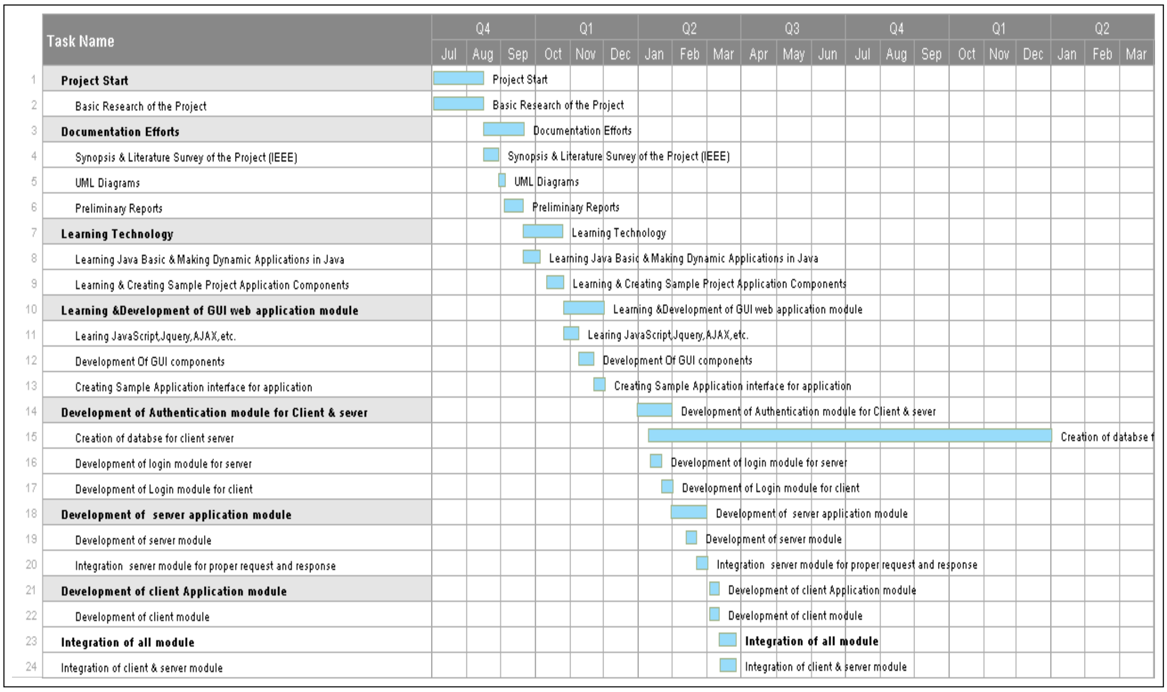


Figure 5.1: Task Network Diagram

### 5.3.3 Timeline Chart

Project planning is part of project management, which relates to the use of schedules such as Gantt charts to plan and subsequently report progress within the project environment. A project management plan is the planning document, capturing the entire project end-to-end, covering all project phases, from initiation through planning, execution and closure.

Analysis or prototyping should increase in direct proportion with project size and complexity. 20 to 25 % of effort is normally applied to software design.



| | Task Name |
|---|---|
| 1 | **Project Start** |
| 2 | Basic Research of the Project |
| 3 | **Documentation Efforts** |
| 4 | Synopsis & Literature Survey of the Project (IEEE) |
| 5 | UML Diagrams |
| 6 | Preliminary Reports |
| 7 | **Learning Technology** |
| 8 | Learning Java Basic & Making Dynamic Applications in Java |
| 9 | Learning & Creating Sample Project Application Components |
| 10 | **Learning &Development of GUI web application module** |
| 11 | Learing JavaScript,Jquery,AJAX,etc. |
| 12 | Development Of GUI components |
| 13 | Creating Sample Application interface for application |
| 14 | **Development of Authentication module for Client & sever** |
| 15 | Creation of databse for client server |
| 16 | Development of login module for server |
| 17 | Development of Login module for client |
| 18 | **Development of server application module** |
| 19 | Development of server module |
| 20 | Integration server module for proper request and response |
| 21 | **Development of client Application module** |
| 22 | Development of client module |
| 23 | **Integration of all module** |
| 24 | Integration of client & server module |

Figure 5.2: Timeline Chart

1. Requirement gathering
2. Literature Survey of existing systems
3. Requirement Modeling and training
4. Development of mock screens
5. Actual Implementation

The Gantt chart for the project is drawn below. The Gantt chart shows the project planning right from the beginning when the topic was finalized.

It depicts the software development life cycle (SDLC). The milestones in the project include topic selection, requirements gathering, Software Requirements Specification, Hardware Requirements Specification.

The milestones also depict the project planning stage. In the Gantt chart below the milestones are represented according to the months in the development lifetime.

## 5.4 TEAM ORGANIZATION

The team for B.E. final year project consists of a team of college students, a college professor as an internal guide and industry professionals as external guide making collaborative efforts for fulfillment and implementation of project problem statement

### 5.4.1 Team Structure

Each and every member of the team is responsible for the identification of problems, proposing problem solving methodologies, identifying approaches for implementation and documentation.

| Sr. No. | Member | Responsibilities |
|---|---|---|
| 1 | Gauri Kavitkar | Project analysis,Developer and Design |
| 2 | Sarvesh Kher | Requirement Gathering And Developer |
| 3 | Gaurav Mahendrakar | Requirement Gathering And Developer |
| 4 | Mayuresh Patil | Testing and Design |

Ms.Nikhita Nerkar is the internal college guide for providing thorough domain guidance, doubt removal and suggesting approaches and ensuring timely completion of activities.

### 5.4.2 Management Reporting And Communication

We report the progress of our project to our internal guide twice a week. We show our weekly status to our guide and incorporate the necessary changes. We communicate among ourselves in case we want suggestions while executing our tasks.

# Chapter 6

# PROJECT IMPLEMENTATION

The systems GUI was designed using java JSP. Core Technologies used were Java, JSP. The overall development was done in the Eclipse 3.3 Indigo and for DB we used MY SQL GUI browser.

## 6.1 OVERVIEW OF PROJECT MODULES

Th eproposed searchable Encryption provides security and secures data when outsourcing to cloud. The important modules used in the project are as follows:
1. Store documents on cloud
2. Key Exchange with User
3. Key Management (ECC Algorithm)
4. Multi-party searchable encryption
5. Content Based Filtering using TF-IDF

## 6.2 TOOLS AND TECHNOLOGIES USED

**JDK 1.7:**

1. The Java Development Kit (JDK) is a software development environment used for developing Java applications and applets. It includes the Java Run- time Environment (JRE), an interpreter/loader (java), a compiler (javac), an archiver

(jar), a documentation generator (javadoc) and other tools needed in Java development.

2. A Java virtual machine (JVM) is an abstract computing machine that enables a computer to run a Java program. There are three notions of the JVM: specification, implementation, and instance. The specification is a document that formally describes what is required of a JVM implementation. Having a single specification ensures all implementations are inter-operable.

3. A JVM implementation is a computer program that meets the requirements of the JVM specification. An instance of a JVM is an implementation running in a process that executes a computer program compiled into Java bytecode.

**Databases**

The database basically used for user storing user details like Username and Password. The tool used for db functionalities was MYSQL GUI Browser.

## 6.3 ALGORITHM DETAILS

### 6.3.1 AES Algorithm

- AES is based on a design principle known as a substitution-permutation network, and is fast in both software and hardware. Unlike its predecessor DES, AES does not use a Feistel network. AES is a variant of Rijndael which has a fixed block size of 128 bits, and a key size of 128 bits. By contrast, the Rijndael specification per se is specified with block and key sizes that may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits.

- AES operates on a 44 column-major order matrix of bytes, termed the state, although some versions of Rijndael have a larger block size and have additional columns in the state. Most AES calculations are done in a special finite field.

- The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the cipher text. The number of cycles of repetition are as follows:

- 10 cycles of repetition for 128-bit keys.
- Each round consists of several processing steps, each containing four similar but different stages, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform cipher text back into the original plaintext using the same encryption key.

### 6.3.2 ECC Algorithm

- Elliptic-curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptic curves can be used for encryption by combining the key agreement with a symmetric encryption scheme.
- In todays world ECC algorithm is used in case of key exchanges by certificate authority (CA) to share the public key certificates with end users. Elliptic Curve Cryptography is a secure and more efficient encryption algorithm than RSA.
- The security of ECC algorithm depends on its ability to compute a new point on the curve given the product points and encrypt this point as information to be exchanged between the end users.
- The ECC system is based on the concepts of Elliptic Curves. To analyze the time taken by an algorithm researches have introduced polynomial time algorithms and exponential time algorithms. Algorithms with smaller computation can be evaluated with polynomial time algorithms and complex computations can be evaluated with exponential time algorithms. The fig 3 show are simple elliptic curve.
- The equation of an elliptic curve is given as, y2 = x3 + ax + b

#### a. Key generation

Key generation is an important part where an algorithm should generate both public key and private key. The sender will be encrypting the message with receivers public key and the receiver will decrypt its private key. Now, select a number, d within the

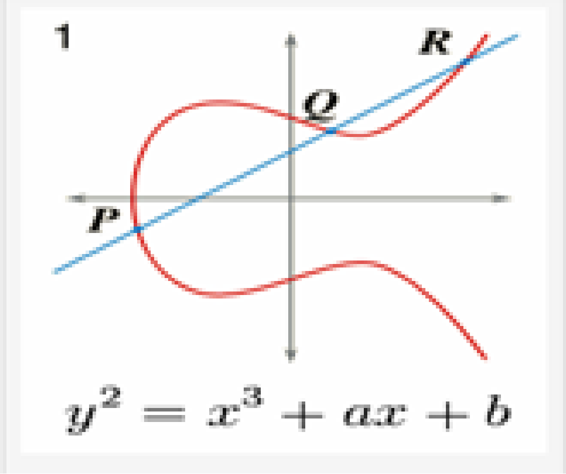


Figure 6.1: Simple Elliptic Curve

range of n. Generate the public key using the following equation,

$$Q = d * P$$

Where d = the random number selected within the range of (1to n-1). P is the point on the curve, Q is the public key and d is the private key.

**b. Encryption**

Let m be the message that has to be sent. Consider m has the point M on the curve E. Randomly select k from [1 -(n-1)]. Two cipher texts will be generated let it be C1 and C2.

C1 = k * P

C2 = M + (k * P)

**c. Decryption**

Use the following equation to get back the original message m that was sent.

$$M = C2 - d * C1$$

M is the original message that was sent.

### 6.3.3 TF-IDF Algorithm

- TF*IDF is an information retrieval technique that weighs a terms frequency (TF) and its inverse document frequency (IDF). Each word or term has its respective TF and IDF score. The product of the TF and IDF scores of a term is called the TF*IDF weight of that term.

- The TF*IDF algorithm is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred to as corpus.

- For a term t in a document d, the weight Wt,d of term t in document d is given by: Wt,d = TFt,d $log\frac{N}{DFt}$ Where:
  TFt,d is the number of occurrences of t in document d.
  DFt is the number of documents containing the term t.
  N is the total number of documents in the corpus.

### 6.3.4 Secure Hashing Algorithm-1

- In cryptography, SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function designed by the United States National Security Agency and is a U.S. Federal Information Processing Standard published by the United States NIST. SHA-1 produces a 160-bit (20-byte) hash value known as a message digest. A SHA-1 hash value is typically rendered as a hexadecimal number, 40 digits long.

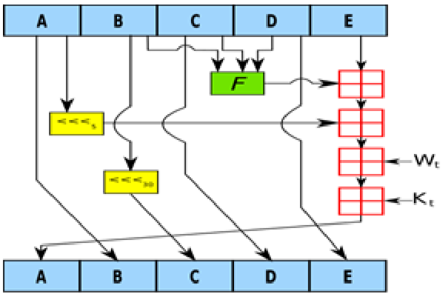


Figure 6.2: Secure Hash Algorithm 1

Image Description: One iteration within the SHA-1 compression function: A, B, C, D and E are 32-bit words of the state; F is a nonlinear function that varies; n denotes a left bit rotation by n places; n varies for each operation; Wt is the expanded message word of round t; Kt is the round constant of round t; denotes addition modulo 232.

# Chapter 7

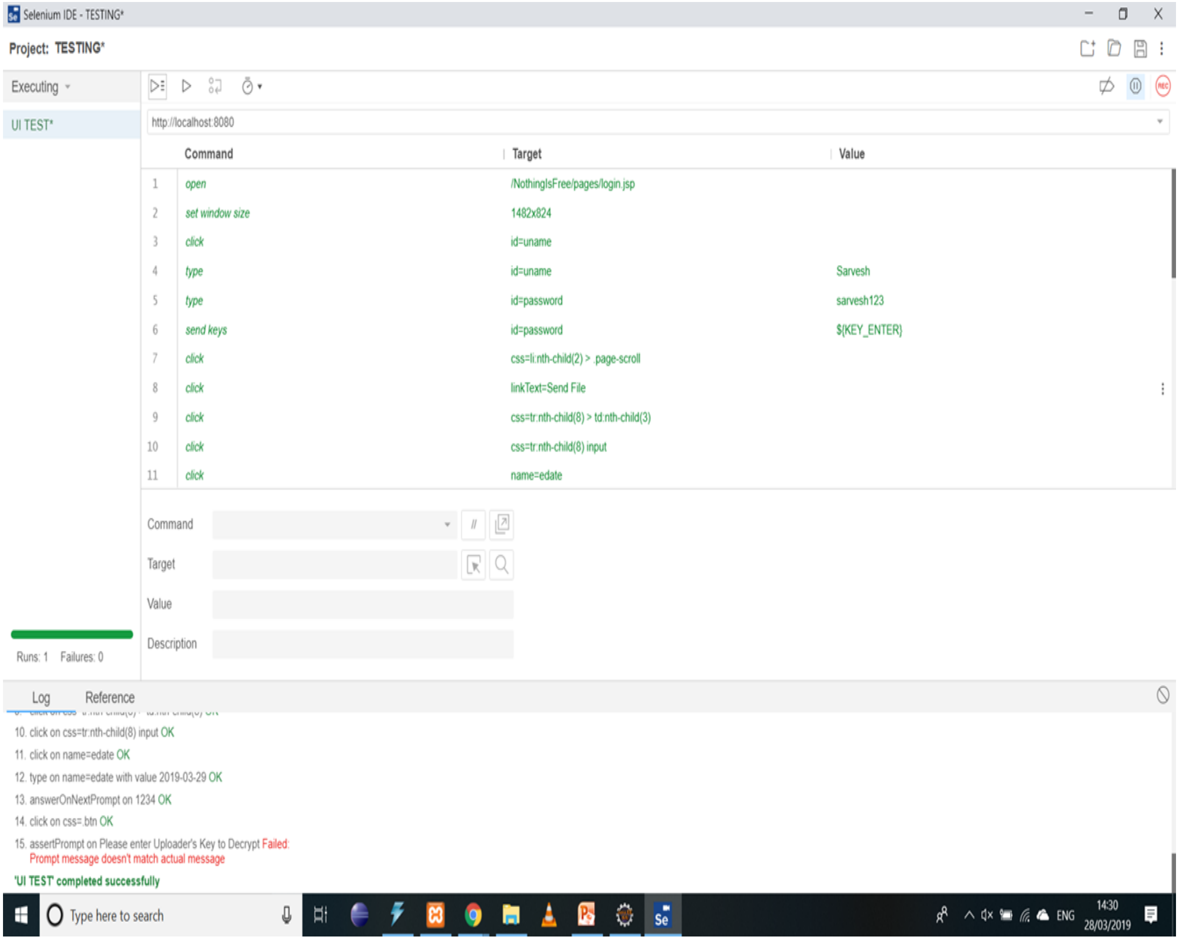
# SOFTWARE TESTING

Software testing is an activity aimed at evaluating an attribute or capability of a program or system and determining that it meets its required results. It is more than just running a program with the intention of finding faults. Every project is new with different parameters. No single yardstick maybe applicable in all circumstances. This is a unique and critical area with altogether different problems. Although critical to software quality and widely deployed by programs and testers. Software testing steel remains an art, due to limited understanding of principles of software. The difficulty stems from complexity of software. The purpose of software testing can be quality assurance, verification and validation or reliability estimation. Testing can be used as a generic metric as well. Software testing is a trade-off between budget, time and quality.

Selenium IDE (Integrated Development Environment) is a prototyping tool for building test scripts. It is a Firefox and Chrome plugin and provides an easy-to-use interface for developing automated tests. Selenium IDE has a recording feature, which records user actions as they are performed and then exports them as a reusable script in one of many programming languages that can be later executed.

## 7.1 TEST CASES & TEST RESULTS

The screenshots of test case is shown in below figure.



| | Command | Target | Value |
|---|---|---|---|
| 1 | open | /NothingIsFree/pages/login.jsp | |
| 2 | set window size | 1482x824 | |
| 3 | click | id=uname | |
| 4 | type | id=uname | Sarvesh |
| 5 | type | id=password | sarvesh123 |
| 6 | send keys | id=password | ${KEY_ENTER} |
| 7 | click | css=li:nth-child(2) > .page-scroll | |
| 8 | click | linkText=Send File | |
| 9 | click | css=tr:nth-child(8) > td:nth-child(3) | |
| 10 | click | css=tr:nth-child(8) input | |
| 11 | click | name=edate | |

```
10. click on css=tr:nth-child(8) input OK
11. click on name=edate OK
12. type on name=edate with value 2019-03-29 OK
13. answerOnNextPrompt on 1234 OK
14. click on css=.btn OK
15. assertPrompt on Please enter Uploader's Key to Decrypt Failed:
    Prompt message doesn't match actual message
'UI TEST' completed successfully
```

Figure 7.1: Test Cases

# Chapter 8

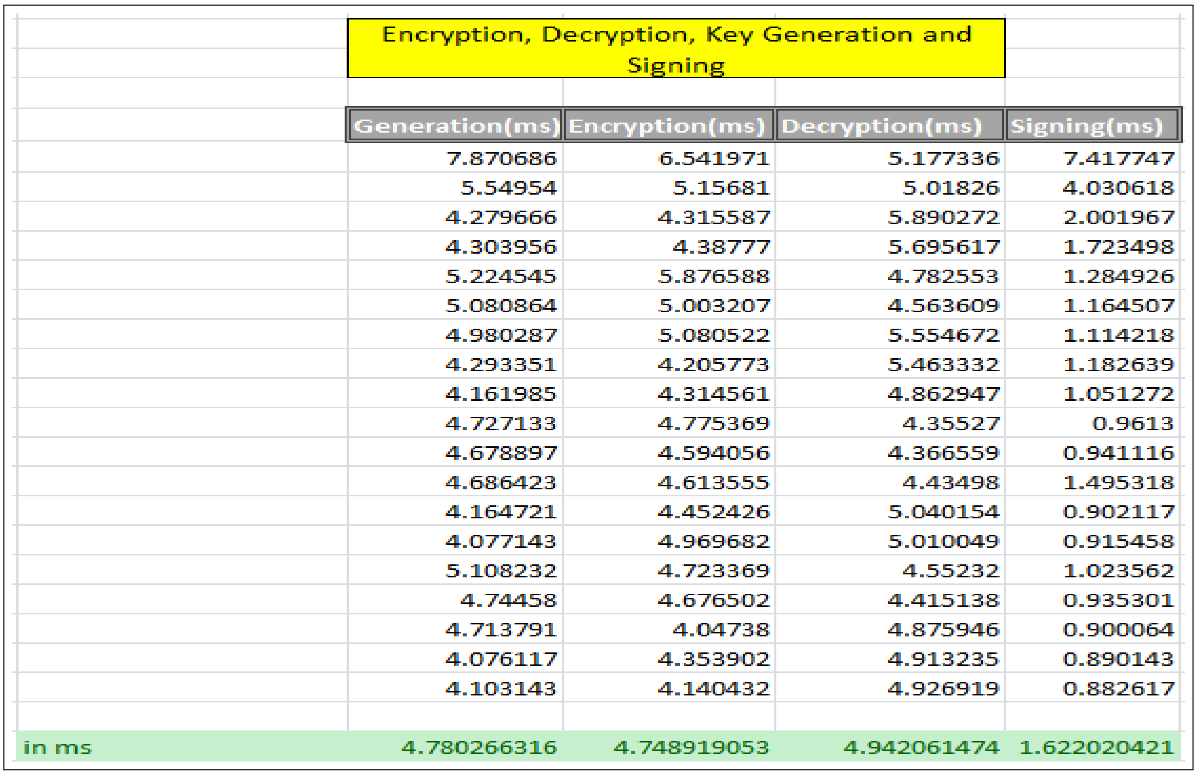
# RESULTS

## 8.1 OUTCOMES

Our solutions use cloud server for encrypted retrieval and make contribution on search accuracy and efficiency. To improve accuracy, we extend the concept hierarchy to expand the search conditions.Authorization information along with index file of each document helps to detect malicious users and prevent them from accessing users private data.Experiments on real world dataset illustrate that our scheme is efficient.
Many systems are proposed to make encrypted data searchable based on keywords. However, keyword-based search schemes ignore the semantic representation information of users retrieval, and cannot completely meet with users search intention.

Selenium IDE (Integrated Development Environment) is a prototyping tool for building test scripts. It is a Firefox and Chrome plugin and provides an easy-to-use interface for developing automated tests. Selenium IDE has a recording feature, which records user actions as they are performed and then exports them as a reusable script in one of many programming languages that can be later executed.

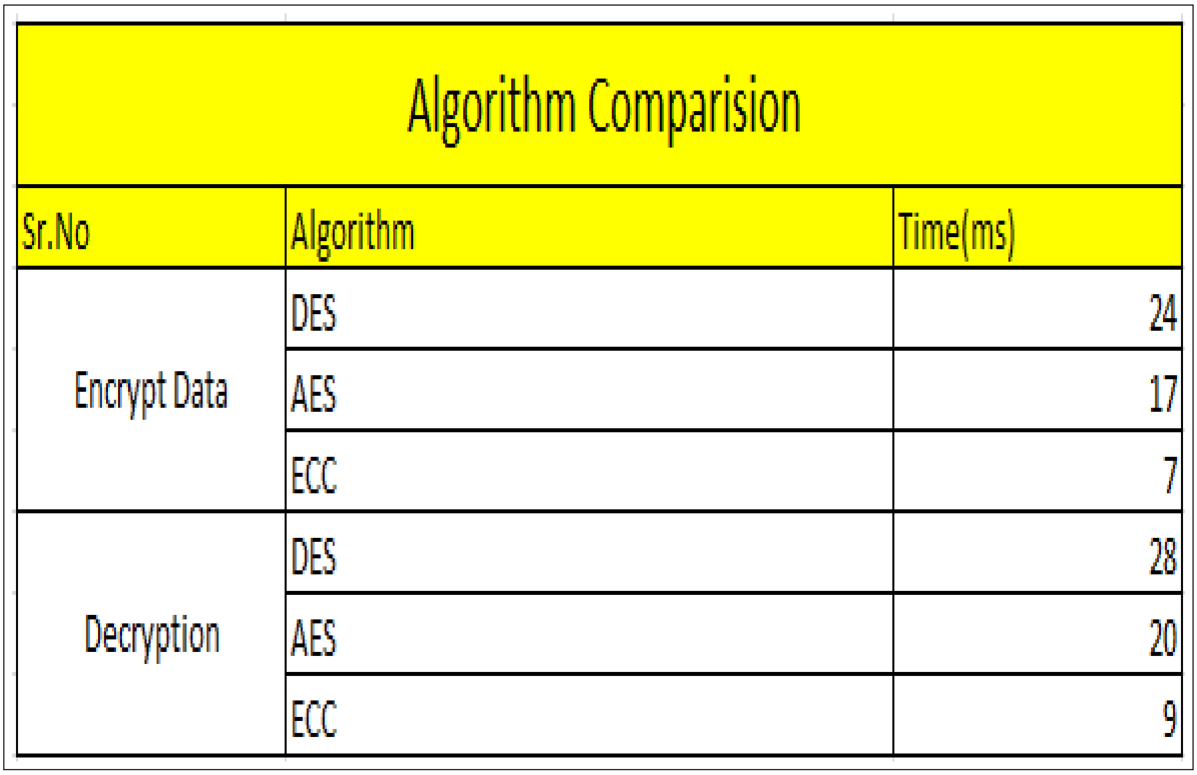
## 8.2 SCREEN SHOTS



Encryption, Decryption, Key Generation and Signing

| | Generation(ms) | Encryption(ms) | Decryption(ms) | Signing(ms) |
|---|---|---|---|---|
| | 7.870686 | 6.541971 | 5.177336 | 7.417747 |
| | 5.54954 | 5.15681 | 5.01826 | 4.030618 |
| | 4.279666 | 4.315587 | 5.890272 | 2.001967 |
| | 4.303956 | 4.38777 | 5.695617 | 1.723498 |
| | 5.224545 | 5.876588 | 4.782553 | 1.284926 |
| | 5.080864 | 5.003207 | 4.563609 | 1.164507 |
| | 4.980287 | 5.080522 | 5.554672 | 1.114218 |
| | 4.293351 | 4.205773 | 5.463332 | 1.182639 |
| | 4.161985 | 4.314561 | 4.862947 | 1.051272 |
| | 4.727133 | 4.775369 | 4.35527 | 0.9613 |
| | 4.678897 | 4.594056 | 4.366559 | 0.941116 |
| | 4.686423 | 4.613555 | 4.43498 | 1.495318 |
| | 4.164721 | 4.452426 | 5.040154 | 0.902117 |
| | 4.077143 | 4.969682 | 5.010049 | 0.915458 |
| | 5.108232 | 4.723369 | 4.55232 | 1.023562 |
| | 4.74458 | 4.676502 | 4.415138 | 0.935301 |
| | 4.713791 | 4.04738 | 4.875946 | 0.900064 |
| | 4.076117 | 4.353902 | 4.913235 | 0.890143 |
| | 4.103143 | 4.140432 | 4.926919 | 0.882617 |
| | | | | |
| in ms | 4.780266316 | 4.748919053 | 4.942061474 | 1.622020421 |

Figure 8.1: Execution Time(ms)



Algorithm Comparision

| Sr.No | Algorithm | Time(ms) |
|---|---|---|
| Encrypt Data | DES | 24 |
| | AES | 17 |
| | ECC | 7 |
| Decryption | DES | 28 |
| | AES | 20 |
| | ECC | 9 |

Figure 8.2: Algorithm comparison

Following are the algorithms that can be used to implement the encryption and decryption and given are the results provided by each specified algorithm that is the time required for each algorithm for its implementation.

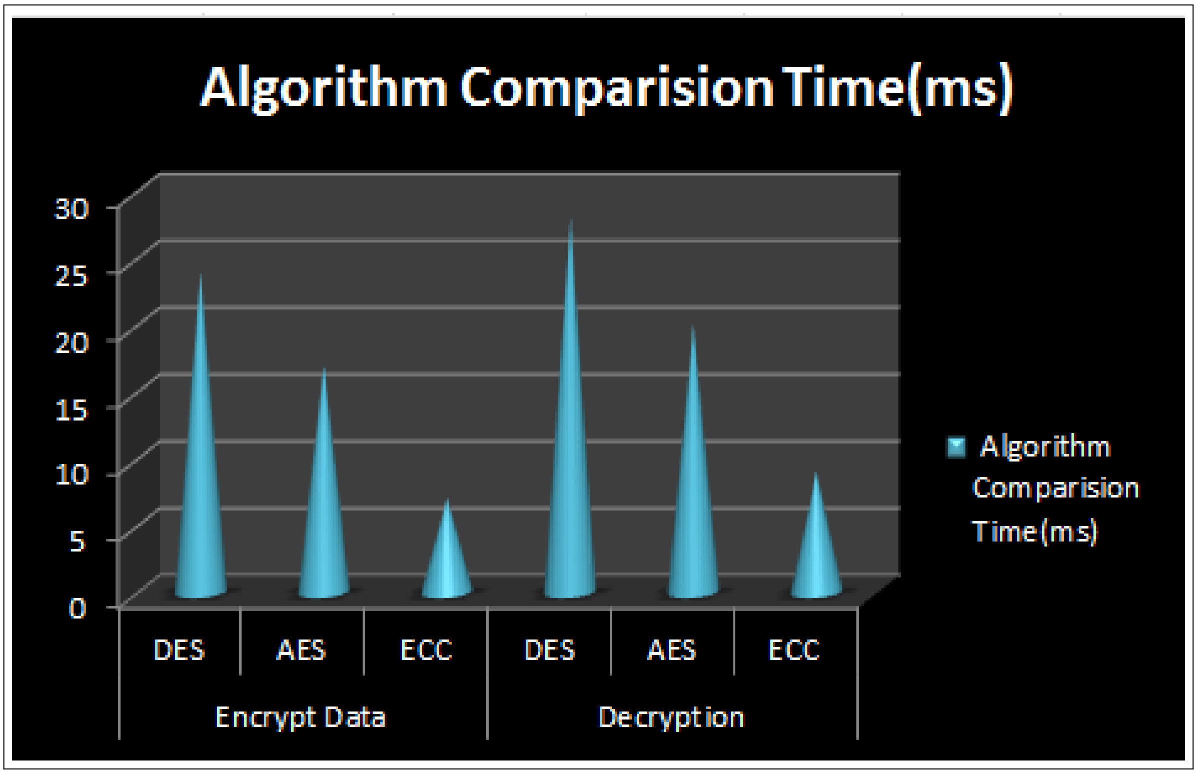


Figure 8.3: Algorithm comparison Time(ms)

# Chapter 9

# CONCLUSION

## 9.1 CONCLUSION

We have formulated a new primitive, namely Semantic-aware Searching over Encrypted Data , for enabling users to selectively authorize each other to search in their encrypted data. Which gives a better improvement than existing system.

When user gets request for searching document at that time he/she sends search index according to the cloud server. Once the data will searched then data will decrypt using AES algorithm. By sign in to the system users's role and department are extracted from login information that is used for extracting users trapdoor key from the key.For searching the document an TF-IDF technique is used which can provide excellent time efficiency.

## 9.2 FUTURE WORK

Many different adaptations, tests, and experiments have been left for the future due to lack of time. Word Net does not include information about etymology or the pronunciation of words and it contains only limited information about usage. Word Net aims to cover most of everyday English and does not include much domain-specific.

## 9.3 APPLICATIONS

Data security Application over cloud

## 9.4 IMPLEMENTATION STEPS

- Open xampp,start apache and mysql server
- Run the code(index.php)
- Localhost8080 : Nothing is free
- Now you are at login page
- Register account User1 and User2
- During registration users will enter their secret keys
- Upload different files on both account
- User 2 will search a word file and from results request for file
- User 1 will give access to read/write and in the backend software will check for semantic relations
- User 2 will decrpyt the file using random key

# Appendix A

# Feasibility Assessment

**Problem statement feasibility assessment using, satisfiability analysis and NP Hard,NP-Complete**

**NP-Hard problem:**

- NP-hard (Non-deterministic Polynomial-time hard), is a class of problems that are, informally, at least as hard as the hardest problems in NP. More precisely, a problem H is NP-hard when every problem L in NP can be reduced in polynomial time.
- As a consequence finding a polynomial algorithm to solve any NP-hard problem would give polynomial algorithms for all the problems in NP, which is unlikely as many of them are considered as hard.
- In TF-IDF we have to compare the each items attributes to rest of the items in the database and a cumulative score is calculated based on their similarities.
- Match these attributes Problems that cannot be solved in fixed time or we can not define their execution complexity with a mathematical algorithm, are called as Non- Deterministic polynomial problems. Therefore, the problem becomes a decision problem, So it is NP.

**NP-COMPLETE:**

- The collection of all problems that can be solved in polynomial time using non-deterministic is called NP. That is, a decision question is in NP if there exists an exponent k and a non-deterministic algorithm for the question that for all hints runs in time O (nk) where n is the length of the input.
- AES is based on a design principle known as a substitution-permutation network, and is fast in both software and hardware.
- AES use 10 cycles of repetition for 128-bit keys. Each round consists of several processing steps, each containing four similar but different stages, including one

that depends on the encryption key itself. All this can be done in polynomial time but requires indefinite time for db interaction so it is NP-COMPLETE.

- All project algorithms can be determined in polynomial time but requires indefinite time for db interaction. Hence all db file handling projects are NP-COMPLETE.

# Appendix B

# Details of Paper Publication

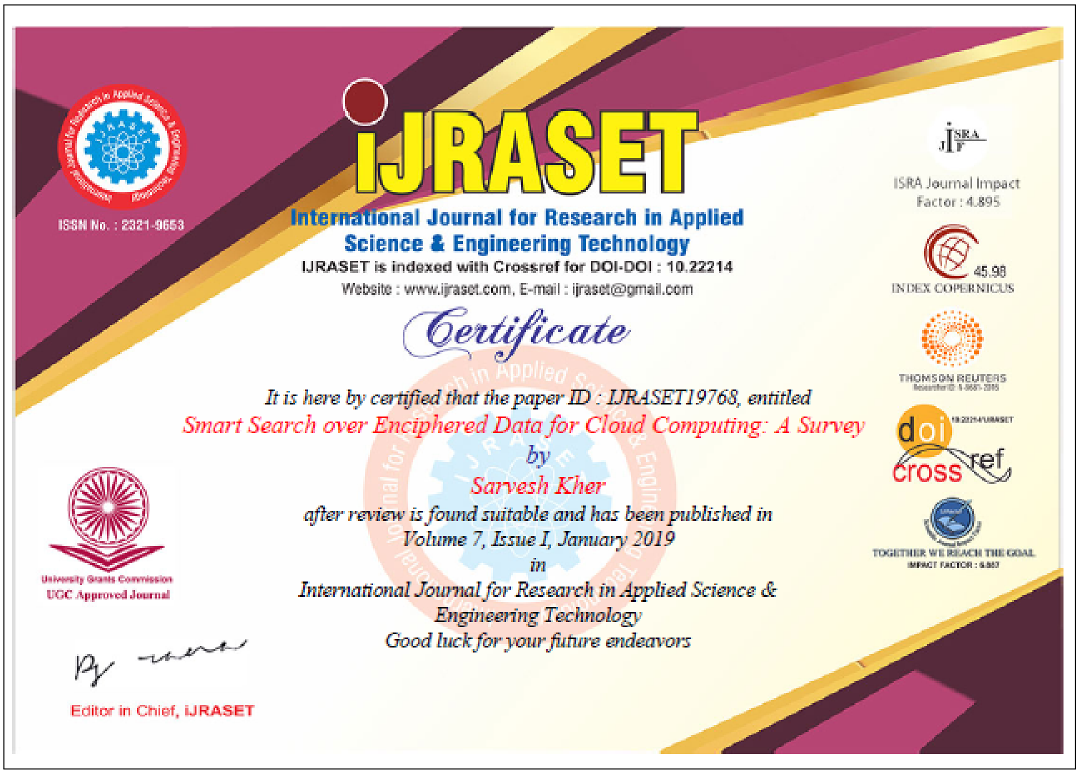
1. **Name of the Conference/Journal**

   **Smart Search over Enciphered Data for Cloud Computing: A Survey** with ID **IJRASET19768,** published in **International Journal for Research in Applied Science and Engineering Technology (IJRASET)**.

2. **Comments of Reviewers**

   Paper Accepted

# Certificate



iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

ISSN No. : 2321-9653

ISRA Journal Impact Factor : 4.895

45.98 INDEX COPERNICUS

THOMSON REUTERS ResearcherID: I-5681-2016

10.22214/IJRASET

TOGETHER WE REACH THE GOAL IMPACT FACTOR : 6.887

Certificate

It is here by certified that the paper ID : IJRASET19768, entitled

Smart Search over Enciphered Data for Cloud Computing: A Survey

by

Sarvesh Kher

after review is found suitable and has been published in

Volume 7, Issue I, January 2019

in

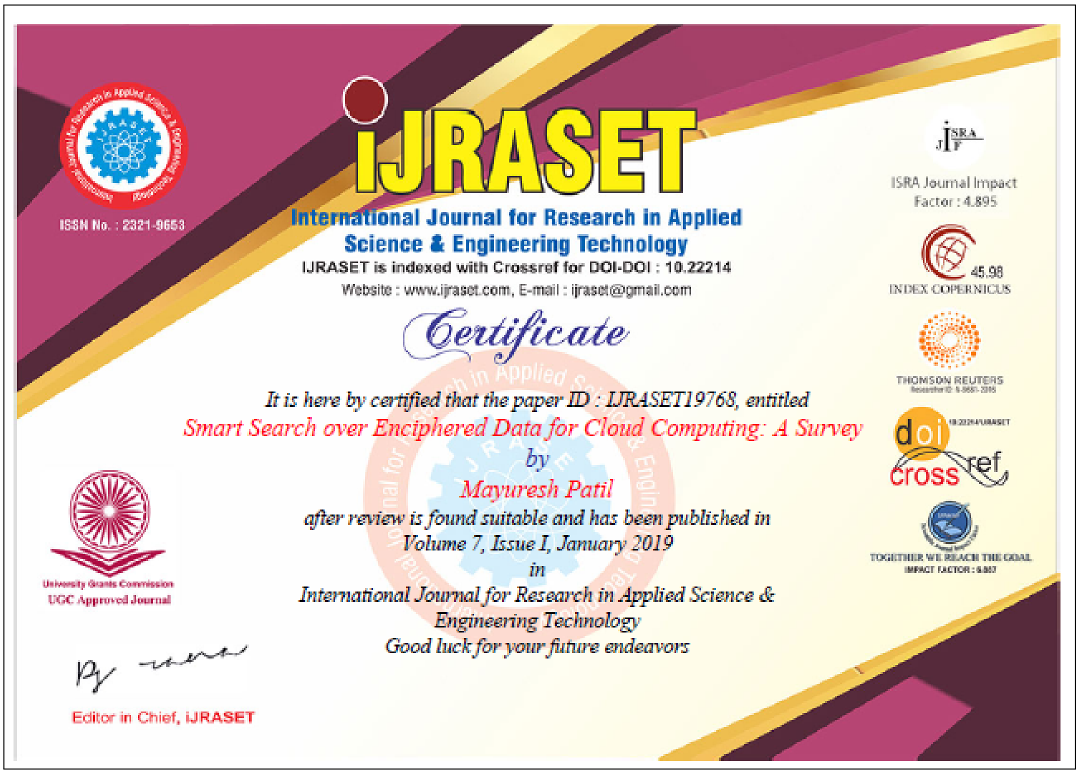
International Journal for Research in Applied Science & Engineering Technology

Good luck for your future endeavors

University Grants Commission UGC Approved Journal

Editor in Chief, iJRASET

Figure 9.1: Certificate



ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

*Certificate*

*It is here by certified that the paper ID : IJRASET19768, entitled*

*Smart Search over Enciphered Data for Cloud Computing: A Survey*

*by*

*Mayuresh Patil*

*after review is found suitable and has been published in*

*Volume 7, Issue I, January 2019*

*in*

*International Journal for Research in Applied Science & Engineering Technology*

*Good luck for your future endeavors*

University Grants Commission
UGC Approved Journal

Editor in Chief, iJRASET

ISRA Journal Impact Factor : 4.895

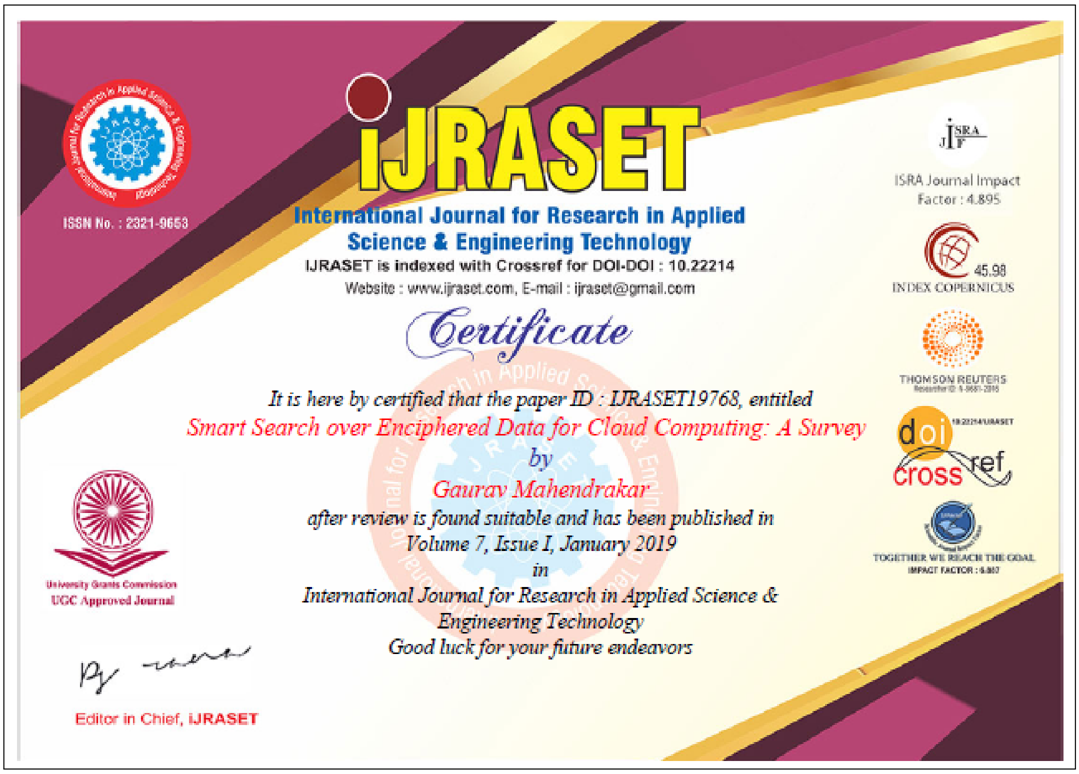
45.98 INDEX COPERNICUS

THOMSON REUTERS
ResearcherID: [illegible]

10.22214/IJRASET

TOGETHER WE REACH THE GOAL
IMPACT FACTOR : 6.887

Figure 9.2: Certificate



ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

Certificate

It is here by certified that the paper ID : IJRASET19768, entitled

Smart Search over Enciphered Data for Cloud Computing: A Survey

by

Gaurav Mahendrakar

after review is found suitable and has been published in

Volume 7, Issue I, January 2019

in

International Journal for Research in Applied Science & Engineering Technology

Good luck for your future endeavors

University Grants Commission
UGC Approved Journal

Editor in Chief, iJRASET

ISRA Journal Impact Factor : 4.895

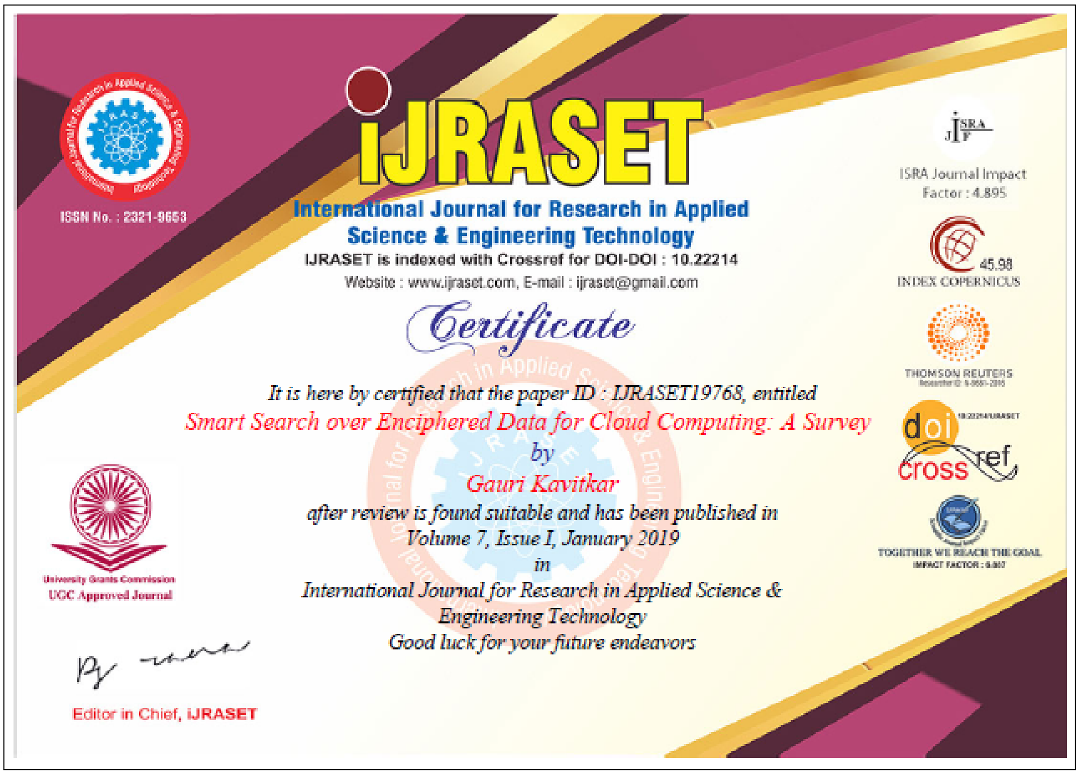
45.98 INDEX COPERNICUS

THOMSON REUTERS

10.22214/IJRASET

TOGETHER WE REACH THE GOAL
IMPACT FACTOR : 6.887

Figure 9.3: Certificate



ISSN No. : 2321-9653

iJRASET

International Journal for Research in Applied Science & Engineering Technology

IJRASET is indexed with Crossref for DOI-DOI : 10.22214

Website : www.ijraset.com, E-mail : ijraset@gmail.com

Certificate

*It is here by certified that the paper ID : IJRASET19768, entitled*

*Smart Search over Enciphered Data for Cloud Computing: A Survey*

*by*

*Gauri Kavitkar*

*after review is found suitable and has been published in*

*Volume 7, Issue I, January 2019*

*in*

*International Journal for Research in Applied Science & Engineering Technology*

*Good luck for your future endeavors*

University Grants Commission
UGC Approved Journal

Editor in Chief, iJRASET

ISRA Journal Impact Factor : 4.895

45.98 INDEX COPERNICUS

THOMSON REUTERS

10.22214/IJRASET

TOGETHER WE REACH THE GOAL
IMPACT FACTOR : 6.887

Figure 9.4: Certificate

# Paper

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

# Smart Search over Enciphered Data for Cloud Computing: A Survey

Sarvesh Kher[1], Mayuresh Patil[2], Gaurav Mahendrakar[3], Gauri Kavitkar[4], Nikhita Nerkar[5]
[1, 2, 3, 4]*Student, Dept. of Computer Engineering, RMD Sinhgad School of Engg, Pune, Maharashtra, India*
[6]*Asst. Professor, Dept. of Computer Engineering, RMD Sinhgad School of Engg, Pune, Maharashtra, India*

***Abstract: With the advancement of information technologies particularly cloud storage used outsourcing data. Now a day's users store a large amount of data on the cloud but it's untrusted and we store secure data on the cloud. The concept of searchable encryption provides a promising direction in solving the privacy problem when** outsourcing data **to the cloud. Such schemes allow users to store their data in an encrypted format an untrusted server and then delegate the server to search on their behalf by issuing a trapdoor (i.e. encrypted keyword). This paper gives a detailed survey of various methods for searchable encryption schemes.***

***Keywords: Encrypted Keyword, AES, Data Outsourcing, Error correction code rule, TF-IDF rule, SHA, Cloud Computing, Smart search over Enciphered Data.***

## I. INTRODUCTION

Now a day's users store an outsized quantity of information on the cloud however it's untrusted and that we store secure data on the cloud. With the advancement of data technologies, notably cloud storage services, info outsourcing and sharing became omnipresent in our life. as an example, a user Alice could store her information at Dropbox and share them together with her friends, within the meanwhile, she can also have access to her friends' information. Thanks to the non-public nature of private information, there's associate inherent would like for a user to by selection shares her information with completely different recipients. I observe, what a user will do is to line some access management policies so have confidence the cloud server (e.g. Dropbox) to enforce them. Sadly, this approach isn't realistic thanks to 2 reasons. One is that the users haven't any suggests that to stop the server from accessing their information. The opposite is that, albeit the server is benign, it should even be forced to share users' information with alternative parties (e.g. by the USA national Act). So it's needed to develop the thought of searchable encoding that provides a promising direction in finding the privacy drawback once outsourcing information to the cloud. Such schemes permit users to store their information in encoding from at Associate in nursing untrusted server so delegate the server to look on their behalf by supplying a personal key and encrypted search index.

## II. LITERATURE REVIEW

In the work of Bao et al. [2], a brand new party, particularly user manager, is introduced into the system, to manage multiple users' search capabilities (e.g. alter them to look every other's data), during this extension, the user manager has to be absolutely sure since it's capable of submitting search queries and decrypting encrypted knowledge. This conflicts with our security criteria (i.e. there mustn't be further TTP involved).

In the work from [3], [4], and [5], the authors have investigated order-preserving encoding, wherever the ciphertexts preserve the order the plaintexts so each entity will perform Associate in Nursing equality comparison. Clearly, these schemes conjointly conflict with our security criteria (i.e. leak marginal data to the server).

Most existing rhombohedral searchable encoding schemes aim at permitting a user to source her encrypted knowledge to a cloud server and delegate the latter to go looking on her behalf. These schemes don't qualify as a secure and ascendable answer for the multiparty setting, wherever users source their encrypted knowledge to a cloud server and by selection authorize one another to go looking. Thanks to the chance that the cloud server might conspire with some malicious users, it's a challenge to own a secure and ascendable multiparty searchable encoding (MPSE) theme. this can be shown by our analysis on the Popa–Zeldovich theme, that says that Associate in Nursing honest user might leak all her search patterns although she shares only 1 of her documents with another malicious user. supported the analysis, the paper [6] presents a replacement security model for MPSE by considering the worst case and average-case situations, that capture totally different server-user collusion potentialities. Then they propose Associate in Nursing MPSE theme by using the additive property of Type-3 pairings and prove its security supported the additive Diffie–Hellman variant and rhombohedra external Diffie–Hellman assumptions within the random oracle model.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue I, Jan 2019- Available at www.ijraset.com*

The [13] author introduces a system that develops associate economical looking formula associated with an economical matching formula within the CSP, so as to forestall, the man within the middle attack and impersonation attack. It secures key distribution formula from information the info the information} owner aspect to data user aspect, wherever the information coding keys is distributed in a very secure thanks to the information users.

The [14] paper proposes a completely unique secure search protocol permits that permits that enables completely different completely different knowledge house owners to code the files and indexes with different keys then construct a tree-based index structure for every knowledge owner and code with AOPPF and allows the cloud server to merge encrypted indexes while not knowing any info.

The [7] paper discuss the matter of privacy-preserving top-k keyword similarity search over outsourced cloud knowledge. Taking edit distance as a live of similarity, we tend to initial build up the similarity keyword sets for all the keywords within the knowledge assortment. We tend to then calculate the relevancy several the weather within the similarity keyword sets by the wide used TF-IDF theory. Leverage each the similarity keyword sets and therefore the relevancy scores, we tend to gift a brand new secure and economical tree-based index structure for privacy-preserving top-k keyword similarity search. To forestall potentially applied mathematics attacks, we tend to conjointly introduce a two-server model to separate the association between the index structure and therefore the knowledge assortment in cloud servers. Thorough analysis is given on the validity of search practicality and formal security proofs are bestowed for the privacy guarantee of our resolution. Experimental results on real-world knowledge set more demonstrate the provision and potency of our resolution.

The [12] author introduces a system during which stratified keyword search on remotely keep knowledge is completed by saving files in the cloud and retrieve the files by ransacking through the keywords. It's bestowed in stratified order victimization ranking formula within the index page. Security for knowledge keep in the cloud is completed through saving encrypted files and privacy of knowledge is maintained by providing completely different completely different trapdoors to different users. The stratified analysis is completed by score dynamics

The paper [8] propose a scientific resolution, that refers to as QDMiner, to mechanically mine question aspects by extracting and grouping frequent lists from free text, HTML tags, and repeat regions at intervals prime search results. Experimental results show that an outsized range of lists do exist and helpful question aspects may be deep-mined by QDMiner. They more analyze the matter of list duplication and find higher question aspects may be deep-mined by modeling fine-grained similarities between lists and penalizing the duplicated lists.

The author of [9] paper addresses issue by developing the fine-grained multi-keyword search schemes over encrypted cloud knowledge. They contribute the three-fold. First, they introduce the relevancy scores and preference factors upon keywords that modify the precise keyword search and customized user expertise. Second, they develop a sensible and really efficient multi-keyword search theme.

The planned theme will support difficult logic search the mixed “AND”, “OR” and “NO” operations of keywords. Third, they more use the classier sub-dictionaries technique to attain higher potency on index building, trapdoor generating and question. Lastly, they analyze the protection of the planned schemes in terms of confidentiality of documents, privacy protection of index and trapdoor, and UNLINK ABILITY of the trapdoor. Through in-depth experiments victimization the real-world dataset, they validate the performance of the planned schemes.

In [10] paper, the author considers objects that are labeled with keywords and are embedded in an exceeding vector area. For these datasets, they study queries that provoke the tightest teams of points satisfying a given set of keywords. It proposes a completely unique methodology known as ProMiSH (Projection and Multi-Scale Hashing) that uses random projection and hash-based index structures and achieves high measurability and acceleration. Also, they gift a definite associate degreed an approximate version of the formula. The experimental results on real and artificial datasets show that ProMiSH has up to sixty times of acceleration over progressive tree-based techniques.

The author of [11] paper proposes ECSED; a completely unique linguistics search theme supported the thought hierarchy and also the linguistic relationship between ideas within the encrypted datasets. ECSED uses 2 cloud servers. One is employed to store the outsourced knowledge sets and come to the hierarchic results to data users. The opposite one is employed to cipher the similarity scores between the documents and also the question and send the scores to the first server. To additional improve the search potency; they utilize a tree-based index structure to arrange all the document index vectors. Also, they use the multi-keyword hierarchic search over encrypted cloud knowledge because the basic frame to propose 2 secure schemes. The experiment results based on the $64000 world datasets show that the theme is additional EFFICIENT than previous schemes.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue I, Jan 2019- Available at www.ijraset.com*

## III. ALGORITHM USED

The summary of the various rule employed by the investigator within the previous paper is given below.

*1)* Encryption/Decryption victimization AES rule
*2)* Error correction code rule
*3)* TF-IDF rule
*4)* SHA-1

### *A. Encryption/Decryption victimization AES*

*1)* AES relies on a style principle called a substitution-permutation network and is quick in each computer code and hardware. In contrast to its forerunner DES, AES doesn't use a Feistel network. AES could be a variant of Rijndael that includes a fastened block size of 128 bits, and a key size of 128 bits. in contrast, the Rijndael specification in and of itself is such as with block and key sizes which will be any multiple of thirty two bits, each with a minimum of 128 and a most of 256 bits.
*2)* AES operates on a 4×4 column-major order matrix of bytes termed the state, though some versions of Rijndael have a bigger block size and have further columns within the state. Most AES calculations are exhausted a special finite field.
*3)* The key size used for Associate in Nursing AES cipher specifies the amount of repetitions of transformation rounds that convert the input, known as the plaintext, into the ultimate output, known as the ciphertext. the amount of cycles of repetition is as follows:
*4)* 10 cycles of repetition for 128-bit keys.
*5)* Each spherical consists of many process steps, every containing four similar however completely different stages, as well as one that depends on the coding key itself. a collection of reverse rounds are applied to rework ciphertext into the initial plaintext victimization constant coding key.

### *B. ECC Algorithm*

*1)* Elliptic-curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptic curves can be used for encryption by combining the key agreement with a symmetric encryption scheme.
*2)* In today's world ECC algorithm is used in case of key exchanges by certificate authority (CA) to share the public key certificates with end users. Elliptic Curve Cryptography is a secure and more efficient encryption algorithm than RSA.
*3)* The security of ECC algorithm depends on its ability to compute a new point on the curve given the product points and encrypt this point as information to be exchanged between the end users.

The ECC system is based on the concepts of Elliptic Curves. To analyze the time taken by an algorithm researches have introduced polynomial time algorithms and exponential time algorithms. Algorithms with smaller computation can be evaluated with polynomial time algorithms and complex computations can be evaluated with exponential time algorithms.

The fig shows a simple elliptic curve.

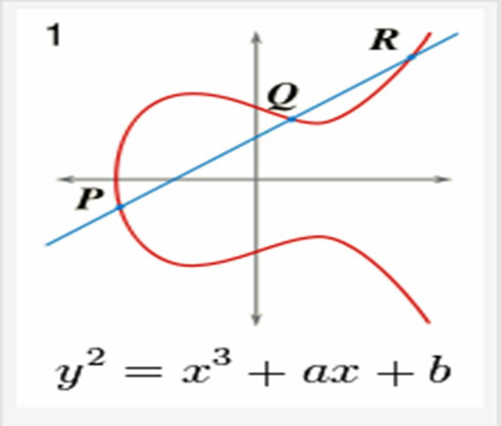


Figure 1: Simple Elliptic Curve.

The equation of an elliptic curve is given as,

$$y2 = x3 + ax + b$$

*a)* *Key Generation:* Key generation is an important part where an algorithm should generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key. Now, select a number, d within the range of n. Generate the public key using the following equation,

$$Q = d * P$$

Where d = the random number selected within the range of (1to n-1). P is the point on the curve, Q is the public key and d is the private key.

*b)* *Encryption*

Use the following equation to get back the original message 'm' that was sent.

$M = C2 - d * C1$

M is the original message that was sent

*c)* *Decryption*

Use the following equation to get back the original message 'm' that was sent.

$M = C2 - d * C1$

M IS THE ORIGINAL MESSAGE THAT WAS SENT

### *C. TF-IDF Algorithm*

*1)* TF*IDF is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term has its respective TF and IDF score. The product of the TF and IDF scores of a term is called the TF*IDF weight of that term.

*2)* The TF*IDF algorithm is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred to as corpus.

For a term t in document d, the weight Wt,d of term t in document d is given by:

$$Wt,d = TFt,d \log (N/DFt)$$

Where:

*a)* TFt,d is the number of occurrences of t in document d.

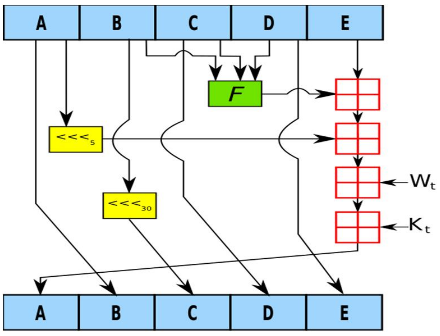
*b)* DFt is the number of documents containing the term t.

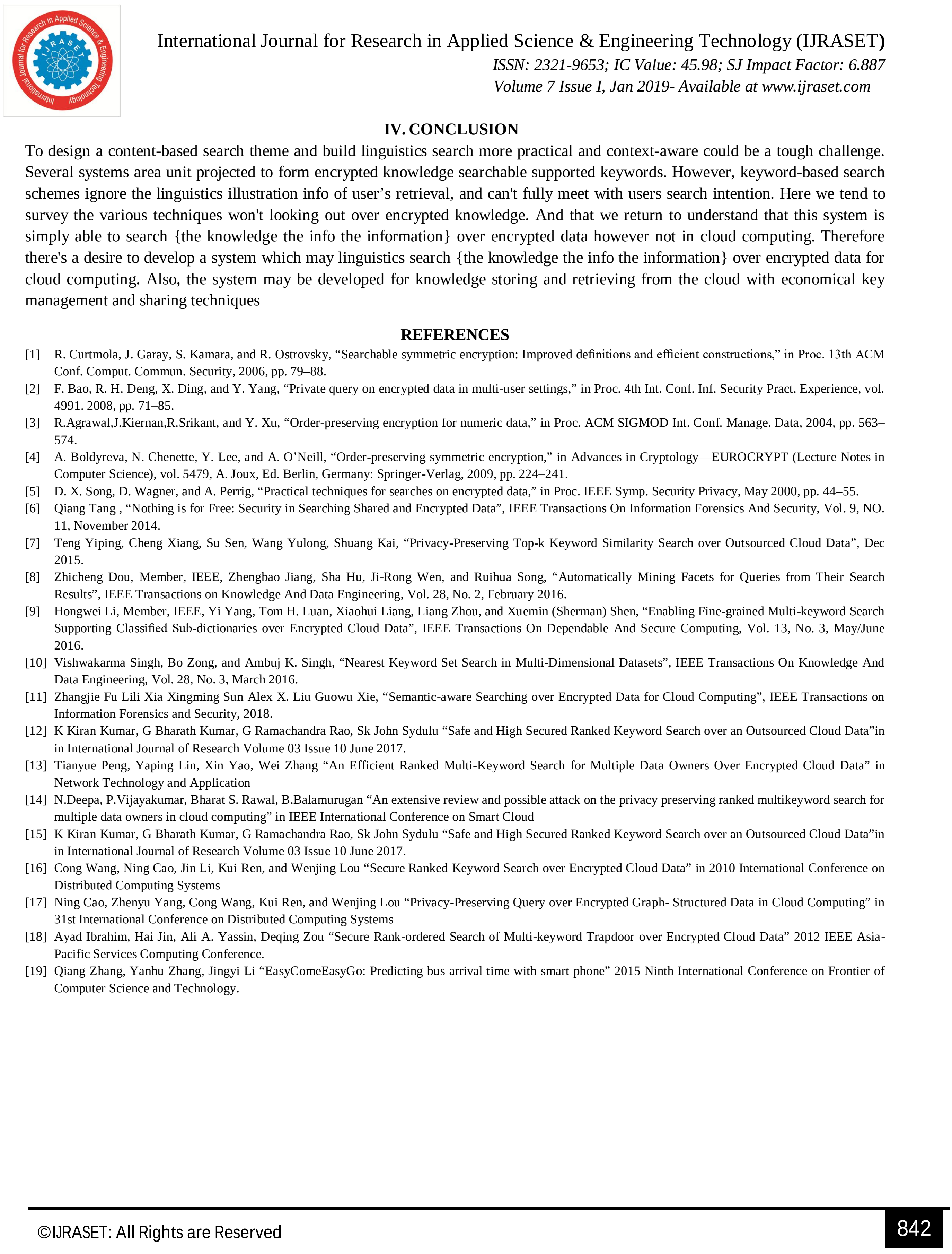
*c)* N is the total number of documents in the corpus.

### *D. Secure Hashing Algorithm-1*

In cryptography, SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function designed by the United States National Security Agency and is a U.S. Federal Information Processing Standard published by the United States NIST. SHA-1 produces a 160-bit (20-byte) hash value known as a message digest. A SHA-1 hash value is typically rendered as a hexadecimal number, 40 digits long

*1)* *Image Description:* One iteration within the SHA-1 compression function: A, B, C, D and E are 32-bit words of the state; *F* is a nonlinear function that varies; *n* denotes a left bit rotation by *n* places; *n* varies for each operation; Wt is the expanded message word of round t; Kt is the round constant of round t; denotes addition modulo 232.





International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887*
*Volume 7 Issue I, Jan 2019- Available at www.ijraset.com*

## IV. CONCLUSION

To design a content-based search theme and build linguistics search more practical and context-aware could be a tough challenge. Several systems area unit projected to form encrypted knowledge searchable supported keywords. However, keyword-based search schemes ignore the linguistics illustration info of user's retrieval, and can't fully meet with users search intention. Here we tend to survey the various techniques won't looking out over encrypted knowledge. And that we return to understand that this system is simply able to search {the knowledge the info the information} over encrypted data however not in cloud computing. Therefore there's a desire to develop a system which may linguistics search {the knowledge the info the information} over encrypted data for cloud computing. Also, the system may be developed for knowledge storing and retrieving from the cloud with economical key management and sharing techniques

## REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 79–88.

[2] F. Bao, R. H. Deng, X. Ding, and Y. Yang, "Private query on encrypted data in multi-user settings," in Proc. 4th Int. Conf. Inf. Security Pract. Experience, vol. 4991. 2008, pp. 71–85.

[3] R.Agrawal,J.Kiernan,R.Srikant, and Y. Xu, "Order-preserving encryption for numeric data," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.

[4] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Order-preserving symmetric encryption," in Advances in Cryptology—EUROCRYPT (Lecture Notes in Computer Science), vol. 5479, A. Joux, Ed. Berlin, Germany: Springer-Verlag, 2009, pp. 224–241.

[5] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Privacy, May 2000, pp. 44–55.

[6] Qiang Tang , "Nothing is for Free: Security in Searching Shared and Encrypted Data", IEEE Transactions On Information Forensics And Security, Vol. 9, NO. 11, November 2014.

[7] Teng Yiping, Cheng Xiang, Su Sen, Wang Yulong, Shuang Kai, "Privacy-Preserving Top-k Keyword Similarity Search over Outsourced Cloud Data", Dec 2015.

[8] Zhicheng Dou, Member, IEEE, Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song, "Automatically Mining Facets for Queries from Their Search Results", IEEE Transactions on Knowledge And Data Engineering, Vol. 28, No. 2, February 2016.

[9] Hongwei Li, Member, IEEE, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, and Xuemin (Sherman) Shen, "Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data", IEEE Transactions On Dependable And Secure Computing, Vol. 13, No. 3, May/June 2016.

[10] Vishwakarma Singh, Bo Zong, and Ambuj K. Singh, "Nearest Keyword Set Search in Multi-Dimensional Datasets", IEEE Transactions On Knowledge And Data Engineering, Vol. 28, No. 3, March 2016.

[11] Zhangjie Fu Lili Xia Xingming Sun Alex X. Liu Guowu Xie, "Semantic-aware Searching over Encrypted Data for Cloud Computing", IEEE Transactions on Information Forensics and Security, 2018.

[12] K Kiran Kumar, G Bharath Kumar, G Ramachandra Rao, Sk John Sydulu "Safe and High Secured Ranked Keyword Search over an Outsourced Cloud Data"in in International Journal of Research Volume 03 Issue 10 June 2017.

[13] Tianyue Peng, Yaping Lin, Xin Yao, Wei Zhang "An Efficient Ranked Multi-Keyword Search for Multiple Data Owners Over Encrypted Cloud Data" in Network Technology and Application

[14] N.Deepa, P.Vijayakumar, Bharat S. Rawal, B.Balamurugan "An extensive review and possible attack on the privacy preserving ranked multikeyword search for multiple data owners in cloud computing" in IEEE International Conference on Smart Cloud

[15] K Kiran Kumar, G Bharath Kumar, G Ramachandra Rao, Sk John Sydulu "Safe and High Secured Ranked Keyword Search over an Outsourced Cloud Data"in in International Journal of Research Volume 03 Issue 10 June 2017.

[16] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou "Secure Ranked Keyword Search over Encrypted Cloud Data" in 2010 International Conference on Distributed Computing Systems

[17] Ning Cao, Zhenyu Yang, Cong Wang, Kui Ren, and Wenjing Lou "Privacy-Preserving Query over Encrypted Graph- Structured Data in Cloud Computing" in 31st International Conference on Distributed Computing Systems

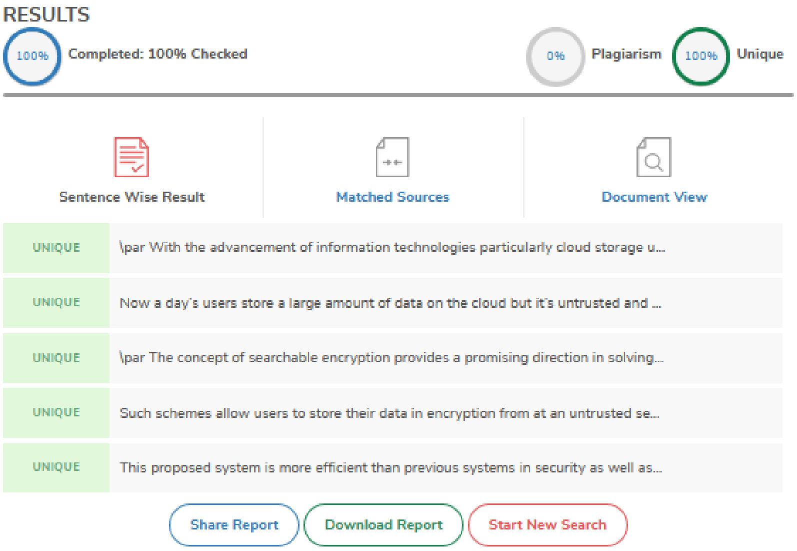
[18] Ayad Ibrahim, Hai Jin, Ali A. Yassin, Deqing Zou "Secure Rank-ordered Search of Multi-keyword Trapdoor over Encrypted Cloud Data" 2012 IEEE Asia-Pacific Services Computing Conference.

[19] Qiang Zhang, Yanhu Zhang, Jingyi Li "EasyComeEasyGo: Predicting bus arrival time with smart phone" 2015 Ninth International Conference on Frontier of Computer Science and Technology.

# Appendix C

# Plagiarism Report

## Plagiarism Report



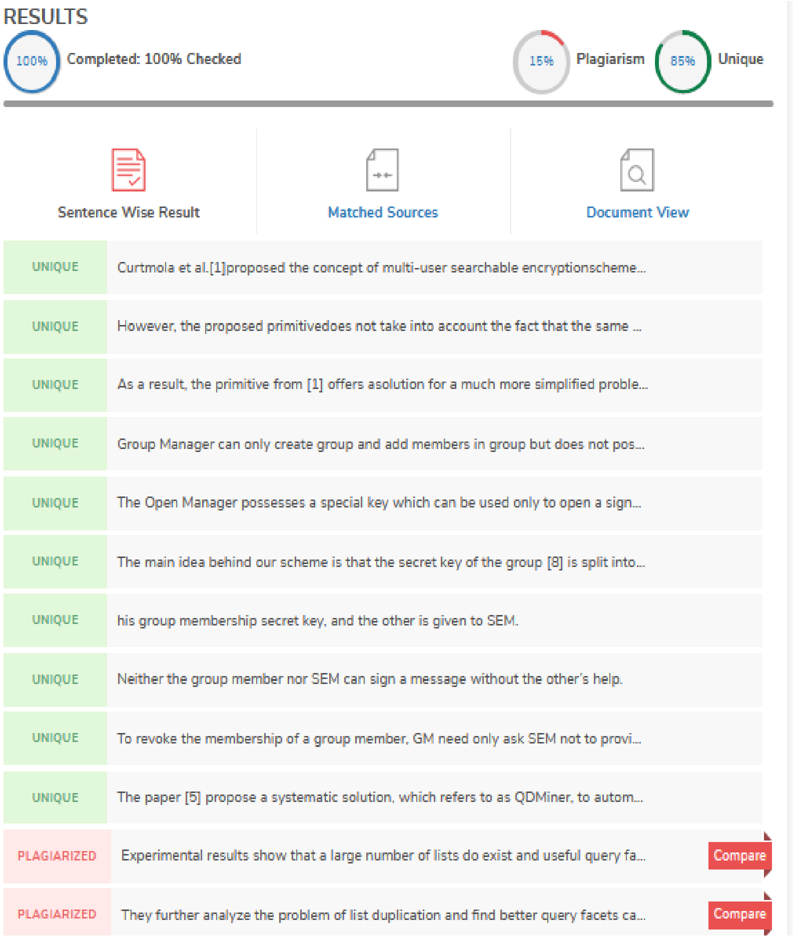
RESULTS

100% Completed: 100% Checked

0% Plagiarism 100% Unique

|  |  |
|---|---|
| UNIQUE | \par With the advancement of information technologies particularly cloud storage u... |
| UNIQUE | Now a day's users store a large amount of data on the cloud but it's untrusted and ... |
| UNIQUE | \par The concept of searchable encryption provides a promising direction in solving... |
| UNIQUE | Such schemes allow users to store their data in encryption from at an untrusted se... |
| UNIQUE | This proposed system is more efficient than previous systems in security as well as... |

Figure 9.5: Abstract Plagiarism



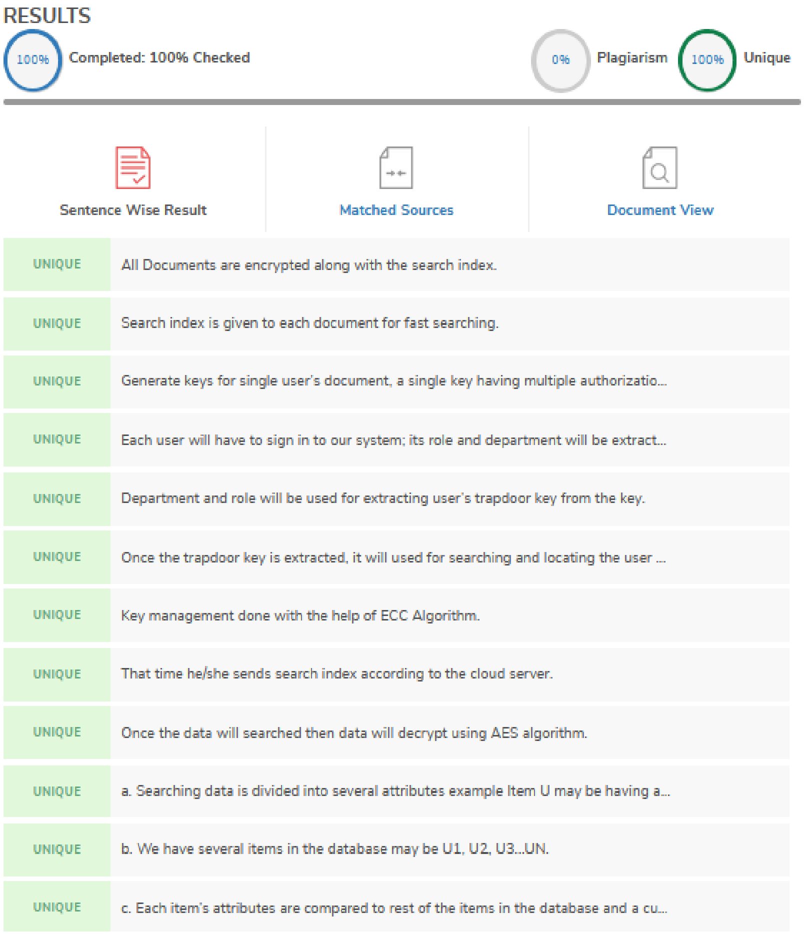
RESULTS

100% Completed: 100% Checked

15% Plagiarism 85% Unique

| | | |
|---|---|---|
| UNIQUE | Curtmola et al.[1]proposed the concept of multi-user searchable encryptionscheme... | |
| UNIQUE | However, the proposed primitivedoes not take into account the fact that the same ... | |
| UNIQUE | As a result, the primitive from [1] offers asolution for a much more simplified proble... | |
| UNIQUE | Group Manager can only create group and add members in group but does not pos... | |
| UNIQUE | The Open Manager possesses a special key which can be used only to open a sign... | |
| UNIQUE | The main idea behind our scheme is that the secret key of the group [8] is split into... | |
| UNIQUE | his group membership secret key, and the other is given to SEM. | |
| UNIQUE | Neither the group member nor SEM can sign a message without the other's help. | |
| UNIQUE | To revoke the membership of a group member, GM need only ask SEM not to provi... | |
| UNIQUE | The paper [5] propose a systematic solution, which refers to as QDMiner, to autom... | |
| PLAGIARIZED | Experimental results show that a large number of lists do exist and useful query fa... | Compare |
| PLAGIARIZED | They further analyze the problem of list duplication and find better query facets ca... | Compare |

Figure 9.6: Literature Suvey Plagiarism



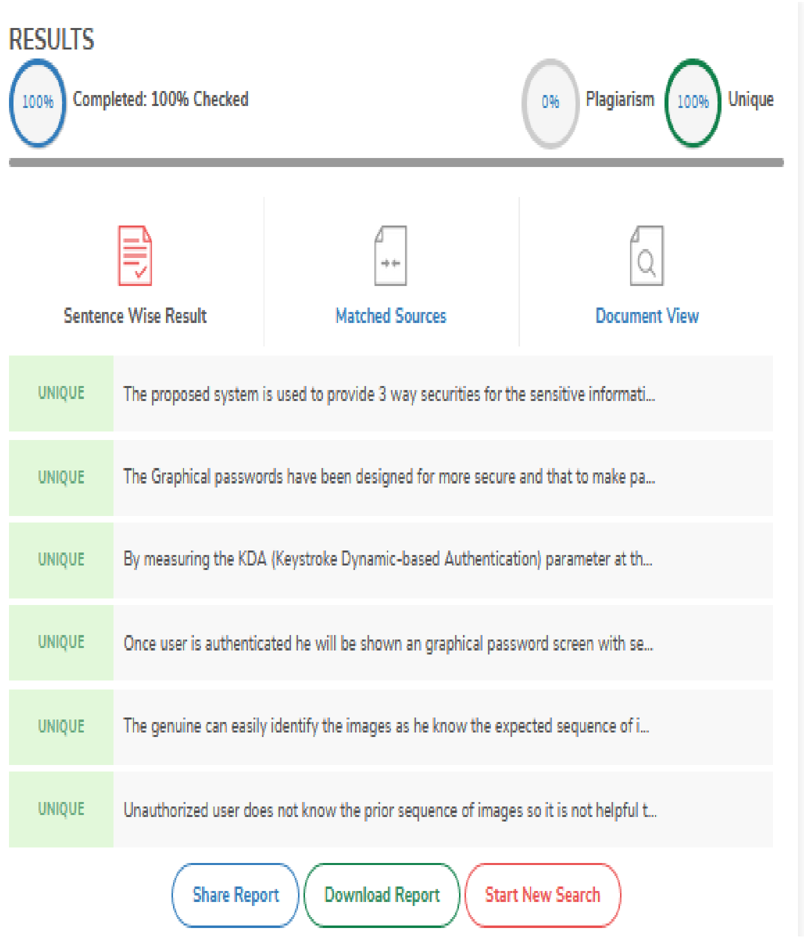
RESULTS

100% Completed: 100% Checked

0% Plagiarism 100% Unique

| | |
|---|---|
| UNIQUE | All Documents are encrypted along with the search index. |
| UNIQUE | Search index is given to each document for fast searching. |
| UNIQUE | Generate keys for single user's document, a single key having multiple authorizatio... |
| UNIQUE | Each user will have to sign in to our system; its role and department will be extract... |
| UNIQUE | Department and role will be used for extracting user's trapdoor key from the key. |
| UNIQUE | Once the trapdoor key is extracted, it will used for searching and locating the user ... |
| UNIQUE | Key management done with the help of ECC Algorithm. |
| UNIQUE | That time he/she sends search index according to the cloud server. |
| UNIQUE | Once the data will searched then data will decrypt using AES algorithm. |
| UNIQUE | a. Searching data is divided into several attributes example Item U may be having a... |
| UNIQUE | b. We have several items in the database may be U1, U2, U3...UN. |
| UNIQUE | c. Each item's attributes are compared to rest of the items in the database and a cu... |

Figure 9.7: Proposed System Plagiarism



| | |
|---|---|
| UNIQUE | The proposed system is used to provide 3 way securities for the sensitive informati... |
| UNIQUE | The Graphical passwords have been designed for more secure and that to make pa... |
| UNIQUE | By measuring the KDA (Keystroke Dynamic-based Authentication) parameter at th... |
| UNIQUE | Once user is authenticated he will be shown an graphical password screen with se... |
| UNIQUE | The genuine can easily identify the images as he know the expected sequence of i... |
| UNIQUE | Unauthorized user does not know the prior sequence of images so it is not helpful t... |

Figure 9.8: Conclusion Plagiarism

# REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, Searchablesymmetric encryption: Improved definitions and efficient constructions, in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 7988.

[2] X. Cao, G. Cong, C. S. Jensen, and B. C. Ooi, Collective spatialkeyword querying, in Proc. ACM SIGMOD Int. Conf. Manage.Data, 2011, pp. 373384. [3]Zhicheng Dou, Member, IEEE, Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song, Automatically Mining Facets for Queries from Their Search Results, IEEE Transactions on Knowledge And Data Engineering, Vol. 28, No. 2, February 2016.

[3] Hongwei Li, Member, IEEE, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, and Xuemin (Sherman) Shen, Enabling Fine-grained Multi-keyword Search Supporting Classied Sub-dictionaries over Encrypted Cloud Data, IEEE Transactions On Dependable And Secure Computing, Vol. 13, No. 3, May/June 2016.

[4] C. Long, R. C.-W. Wong, K. Wang, and A. W.-C. Fu, Collectivespatial keyword queries: A distance owner-driven approach, inProc. ACM SIGMOD Int. Conf. Manage. Data, 2013, pp. 689700.

[5] C. Wang, Q. Wang, K. Ren, and W. Lou, Privacy-preservingpublic auditing for data storage security in cloud computing,inProc. of IEEE INFOCOM 2010, CA, USA, Mar. 2010, pp. 525533.

[6] Pushkar Zagade, Shruti Yadav, Aishwarya Shah, Ravindra Bachate Group User Revocation and Integrity Auditing of Shared Data in Cloud Environment International Journal of Computer Applications (0975 8887) Volume 128 No.12, October 2015.

[7] Subhra Mishra and Tilak Rajan Sahoo A Survey on Group Signature Schemes Department of Computer Science and Engineering National Institute of Technology Rourkela Rourkela.

[8] Tao Jiang, Xiaofeng Chen, and Jianfeng Ma Public Integrity Auditing for Shared Dynamic Cloud Data with Group User Revocation 2015 IEEE.

[9] He Ge An Effective Method to Implement Group Signature with Revocation.

[10] S. Cui, X. Cheng and C. W. Chan, “Practical group signatures from RSA,” 20th International Conference on Advanced Information Networking and Applications - Volume 1 (AINA’06), Vienna, 2006

[11] Vishwakarma Singh, Bo Zong, and Ambuj K. Singh, Nearest Keyword Set Search in Multi-Dimensional Datasets, IEEE Transactions On Knowledge And Data Engineering, Vol. 28, No. 3, March 2016.

[12] Zhangjie Fu Lili Xia Xingming Sun Alex X. Liu Guowu Xie, Semantic-aware Searching over Encrypted Data for Cloud Computing, IEEE Transactions on Information Forensics and Security, 2018.

[13] Sarvesh Kher,Mayuresh Patil,Gaurav Mahendrakar,Gauri Kavitkar and Nikhita Nerkar Smart Search over Enciphered Data for Cloud Computing: A Survey IJRASET.