Privacy-Preserving Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data

Yogesh Singh, Suraj Sonawane, Pratik Bansode, Swapnil Wankhede
BE Students, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India,
Kanchan Jadhav
Professor, Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India.

Abstract- Cloud computing is the on-demand availability of computer system resources, especially data storage and computing power, without direct active management by the user. Cloud computing provides individuals and enterprises massive computing power and scalable storage capacities to support a variety of big data applications in domains like health care and scientific research, therefore more and more data owners are involved to outsource their data on cloud servers for great convenience in data management and mining. Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we present a secure multi-keyword ranked search scheme over encrypted cloud data. We investigate the Multi-keyword top-k search problem for data encryption against privacy breaches, and attempt to identify an efficient and secure solution to this problem. Specifically, for the privacy concern of query data, we construct a database query search method.

Keywords- Multi-Keyword Ranked Search, Encrypted Cloud Data, Public Key Encryption, top-k search.

I. INTRODUCTION

1.1 Overview

Cloud computing is one of the latest developments in the IT industry also known as on demand computing. This technology is grouped into sections which include SAAS, IAAS and PAAS. Now days Cloud computing makes everything flexible and easier but there is another aspect that is what about security? Data security over the Cloud also a major concern and various methodologies are proposed, considering the customer point of view, we have made an extensive research to obtain what are the main security problems in Cloud computing security. Security of the cloud environment depends on the security provided by the cloud service provider. Cloud providers control the hardware and the hyper visors that stores the data and applications are run. First on the list is data breaches. If a Cloud service database isn't designed properly, a single flaw in one client's application could allow an attacker to get at not just that client's data, but every other clients data as well. The challenge in addressing these threats of data loss and data leakage. In this paper we want to develop the security system that will provide security to the cloud and will be very fruitful for both the client user and also the cloud data owner, that we can perform trusted computing. To improve the security for the data retrieval from cloud environment, the Verification Key is used and provides data encryption which protects data from cloud vendor, an attacker. The Verification Key is sent to the user Email Id to view the original data and help to share public data with other authenticated cloud user.

1.2 Background

In today's world, the sharing of data is a vital part of our lives. The data can be sensitive and hence when it gets into the hands of wrong people, it can turn out to be harmful for both the owners and the receivers. Hence to overcome such a situation, encryption and decryption of data can be done for the safe exchange of any amount and type of data. There are two sets of parties which would use this technique; one being the data owners, which would own the data and the next are the data users which do not own the data but use this after getting the permission for the usage. The data owners and the data users need to be authenticated on the cloud servers beforehand to use any of the services provided. When the sensitive data is outsourced to the cloud, so as to enable the easier accessing of the data by the data owners and the data users, it is encrypted. The data encrypted has a list of keywords which are sent to an administration server. This in turn is then re-encrypted and uploaded by the administration server.

When the data users would want to access these encrypted files, they will have to get themselves authenticated. Once the data users are authenticated and verified, they would search the files using keywords. The keywords are sent to the administration server which in turn would encrypt the given keyword. The encrypted keyword is then compared to the existing keywords and the files are given to the data users after the decryption. Hence this helps in creating a secure environment for the exchange of the information among the data owners and the data users.

Multi-Keyword

Multi-keyword refers to the ability of searching multiple keywords at a given time. This means that the user can search n number of keywords at a given time. The file retrieved is the one which either contains all the keywords or a minimum of one of the keywords that the user searches.

Ranked Search

As the user searches for the document over the cloud, like Google, the cloud server can return the document which is the most relevant document amongst the entire collection of documents. The search result can be ranked on different parameters. The parameters can be last downloaded, last visited, recently uploaded, etc. when the user tries to access a document, he/she can specify the k number of documents to be downloaded. When the k is specified, the top k documents are displayed.

Multiple Data Owners

The data owners are the types of users which upload their data over the cloud server. The data owners register themselves at the administrators and hence this gives them the right to upload their data. The data users hence use this data for their future uses.

1.3 Motivation

Cloud security is important for both business and personal users. Everyone wants to know that their information is safe and secure. Businesses have legal obligations to keep client data secure, with certain sectors having more stringent rules about data storage. To prevent unauthorized access to our data we need to provide some security mechanism to our data. Now days the Third-party cloud service providers are increasing very fast rate uploading or using their services may lead to misuse of our data (e.g. Balance sheet, Employee details). To provide security to such important documents and data is our motivation behind this project.

1.4 Problem Definition and Objectives

A general approach to protect the data confidentiality is to encrypt the data before outsourcing. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability. We define and solve the problem of effective yet secure ranked keyword search over encrypted cloud data while also encrypting keyword to be searched for providing more security.

1.5 Project Scope & Limitations

- The proposed system can be used to enhance the security mechanism in FSS (File Sharing Systems).
- The keyword search functionality is performed over encrypted cloud data without leaking any information about the search keywords by encrypting the Keywords.
- We can improve the security mechanism by different dynamic encryption algorithms.
- The files won't be shared to anyone unless and until the user (Data owner) grants the access using security verification key.
- We can improve by encrypting multiple formats of file enhancing Multiple file sharing system.
- There are also, however, limitations when it comes to this technology like large files have problem to be encrypted.

II. RELATED WORK

In the [1] work user suggested, Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. Since regular private-key encryption prevents one from searching over encrypted data, clients also lose the ability to selectively retrieve segments of their data. The area of searchable encryption has been identified by DARPA as one of the technical advances that can be used to balance the need for both privacy and national security in information aggregation systems. In the [2] work the author presented, A related issue deals with privacy of database data. There are two different scenarios: public databases and private databases, and the solutions for each are different. Private databases: In this setting a user wishes to upload its private data to a remote database and wishes to keep the data private from the remote

database administrator. Later, the user must be able to retrieve from the remote database all records that contain a particular keyword.

In the [3] work, the practical considerations and enhancements of our ranked search mechanism, including the efficient support of relevance score dynamics, the authentication of ranked search results are investigated. Some schemes are developed in which user must have knowledge about all the valid keywords and their respective positions as mandatory information so as to generate a query [4]. While web-based applications are gaining popularity, they also pose new security challenges. In particular, recently revealed that many popular Web applications actually leak out highly sensitive data from encrypted traffic due to side-channel attacks using packet sizes and timing [5].

The paper [6] states that using cloud computing, individuals can store their data on remote servers and allow data access to public users through the cloud servers. As the outsourced data are likely to contain sensitive privacy information, they are typically encrypted before uploaded to the cloud. This, however, significantly limits the usability of outsourced data due to the difficulty of searching over the encrypted data. To provide a secure and efficient retrieval of data, one needs to ensure that the user can perform a search over the encrypted data without revealing the information to the server. The cryptographic primitive that provides this feature is widely known as searchable encryption [7]. To maintain data privacy, the keyword search functionality needs to be performed over encrypted cloud data without leaking any information about the search keyword or the retrieved document. This is known as privacy-preserving keyword search [8].

III. SYSTEM ARCHITECTURE

a. Existing system Architecture

A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause tremendous cost in terms of data utility, which makes traditional data processing methods that are designed for plaintext data no longer work well over encrypted data. The keyword-based search is widely used technique in many data storage and retrieval systems like Google search engine and is traditional method can't be applied on encrypted data, then How to process such queries over encrypted data and at the same time maintaining the privacy of user's data becomes challenging task.

b. Proposed System Architecture

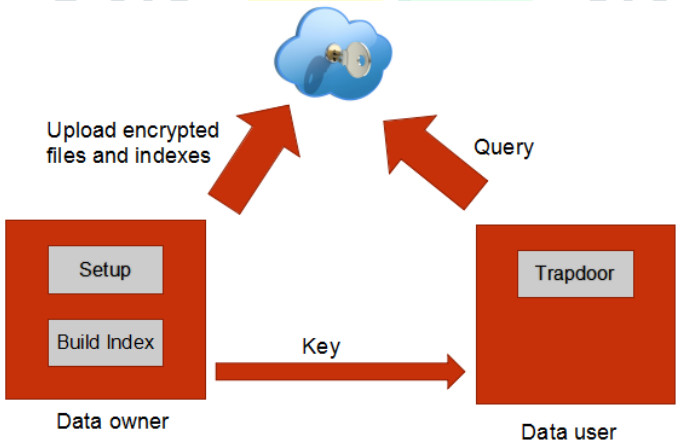


Fig.3.1 Proposed System Architecture

3.1 Overview of Project Modules:

Data Owner upload the detailed information in text file .That file will encrypt with Advanced Encryption Standard algorithm that takes plain text and private key and encrypt data. Encrypted data will store on cloud. Also encrypted keyword will be stored. User will see the file after login and searching with keywords and send request to get the file .Key Manager will send the verification key on authorized mail. Then the user will enter the verification key and will get the decrypted file and can download it.

3.1.1 Owner

- Owner registers to the system.
- Owner login to the system.
- Owner will upload file.
- Owner can view details.
- Owner can access the data.
- Lastly, logout from the system.

3.1.2 CSP

- CSP login to the system.
- CSP can view details.
- CSP can view the data.
- Approve and reject owner and user.
- Lastly, logout from the system.

3.1.3 User

- User registers to the system.
- User login to the system.
- User can view details.
- User can view the data.
- Lastly, logout from the system.

3.2 Mathematical Model

Let us consider S as a system for Secure File Sharing System.
S= I, F, O
INPUT:
Identify the inputs.
F= f1, f2, f3, ..., fn— Set of functions to execute commands.
I= i1, i2— Set of inputs to the function set.
O= o1— Set of outputs from the function sets.

Where,
I = File to be uploaded and keywords given by the user.
O = Users output file.
F = Encryption, Decryption and Searching Functions implemented to get the output.

Space Complexity:
The space complexity depends on size of the File.
More the storage of data more is the space complexity.

Time Complexity:
Check No. of matches available = n
If database is large then retrieving of information can be time consuming.

Failures and Success conditions:
Failures:
Huge database can lead to more time consumption to get the information.
Hardware failure.
Software failure.
Success:
Search the required files using keywords from available in Database.
User gets Top-k ranked search results.

3.3 Algorithm Details

3.3.1 Advanced Encryption Standard:

The more popular and widely adopted symmetric encryption algorithm likely to be encountered nowadays is the Advanced Encryption Standard (AES). It is found at least six times faster than triple DES.
A replacement for DES was needed as its key size was too small. With increasing computing power, it was considered vulnerable against exhaustive key search attack. Triple DES was designed to overcome this drawback but it was found slow.

The features of AES are as follows −

Symmetric key symmetric block cipher.
128-bit data, 128/192/256-bit keys.
Stronger and faster than Triple-DES.

Provide full specification and design details.
Software implementable in C and Java.

3.3.2 Operation of AES

AES Algorithm for Encryption.
Input:
128 bit plain text.
128/192/256-bit keys.
Process:
10/12/14-rounds for-128_bit /192 bit/256 bit key size.
XOR state block (i/p)
Final round:10,12,14
Each round consists: sub byte, shift byte, mix columns, add round key.
Output:
Cipher text.

IV. EXPERIMENTAL SET UP / RESULTS

The application will be able to secure sharing of the data using the dynamically generated Verification Key, by implementing this project we are creating a way to share a data whenever it is approved by the user only, if the data is tried to hack by any service provider then he will get the encoded data that will be required the decryption using 256 bit key. The fastest supercomputer in the world, it will take millions of years to crack 256-bit AES encryption.

In our experimental setup, in table 1, find out number of file upload and file download. In our experimental setup in our system number file upload and download of files.

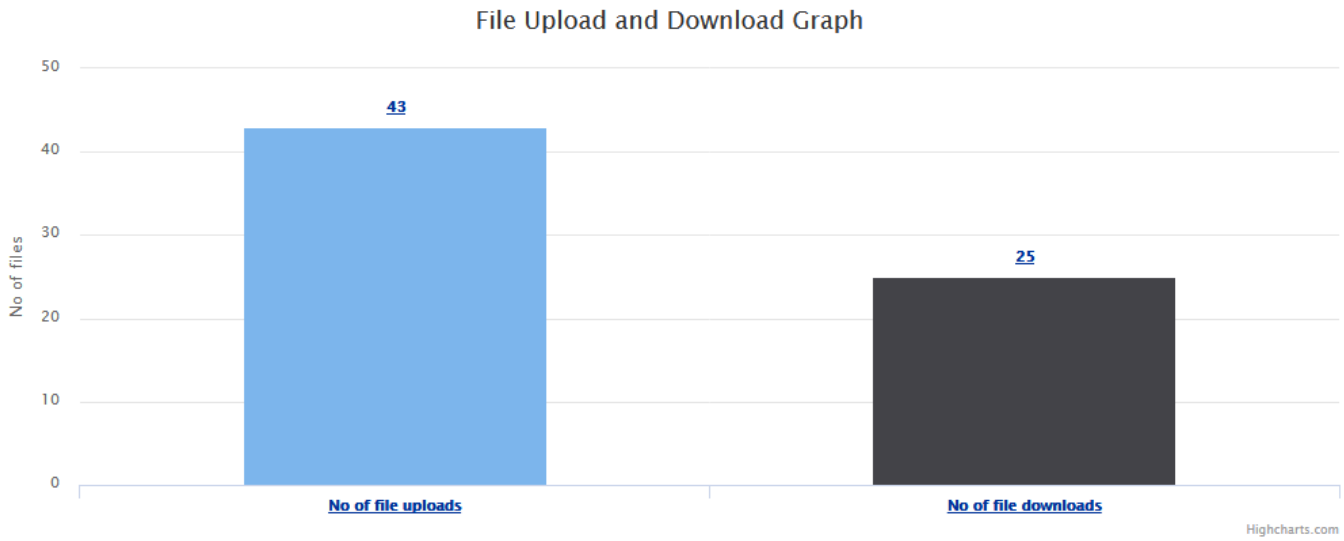
Sr. No	Number of File Upload	Number of File Download
1	43	25

Table 1: No. Upload and download files

In table 2, User can search different file with different keywords so get information about how to user can search any files.

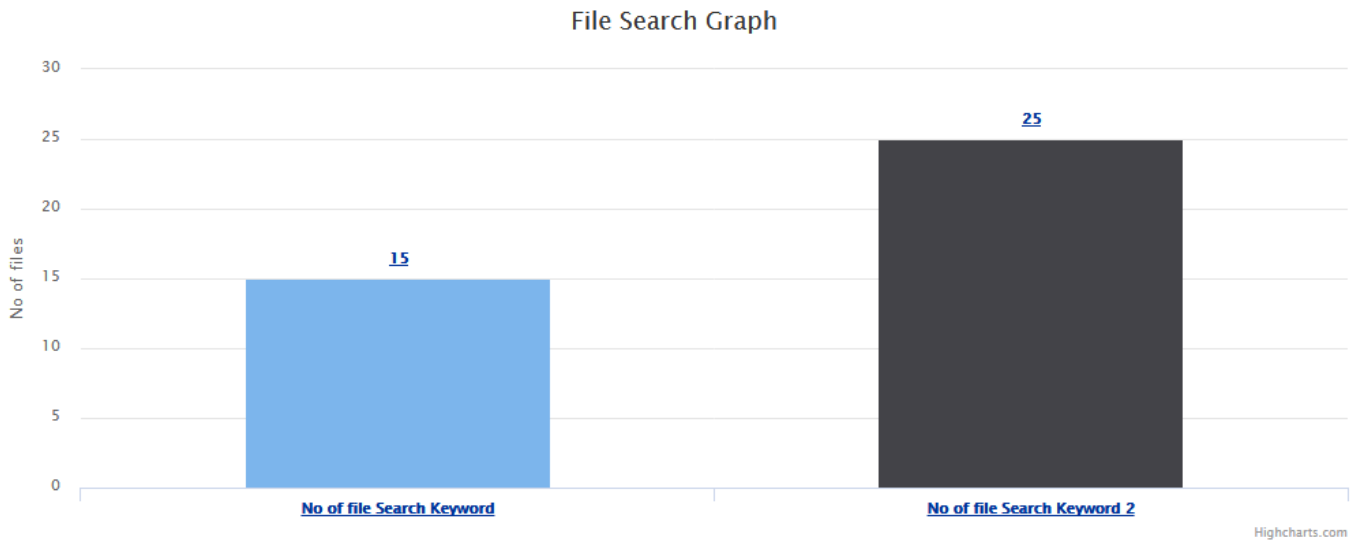
Sr. No	No. of File Match with keyword 1	No. of File Match with keyword2
1	15	25

Table 2: No. File Search keyword



Graph 1: File upload and download graph

From above data, In graph 1, we can see the no. of file upload and no of file download in the graph; we see 43 files upload by different data owners and 25 different users are download in the graph.



Graph 2: File Search graph by different keywords

From above data, in graph 1, we can see the no. of file search keyword and no of file keyword 2 in the graph; we see 15 files searched by keyword 1 and 25 files searched by keyword2 by different users are shown in the graph.

V. SCREENSHOTS OF IMPLEMENTED PROJECT

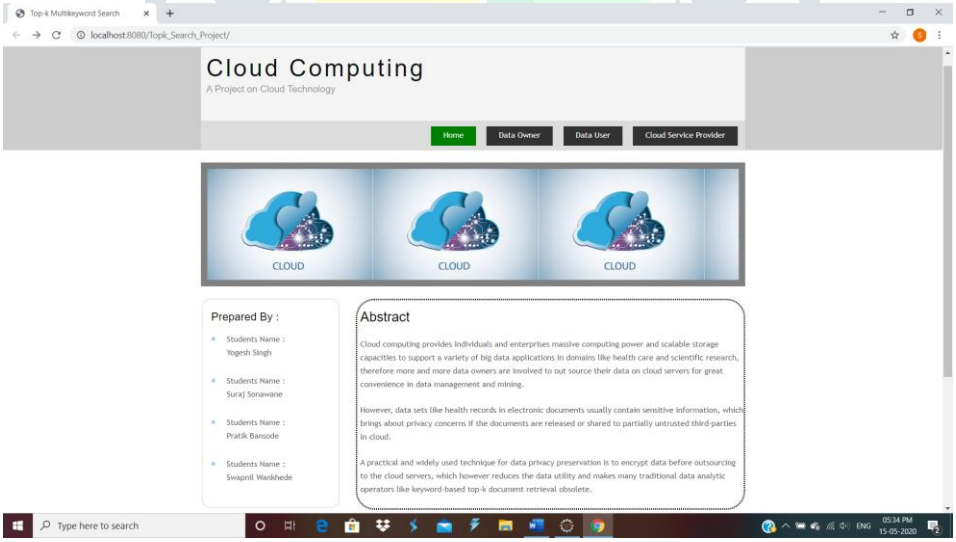


Fig.5.1 Home Page

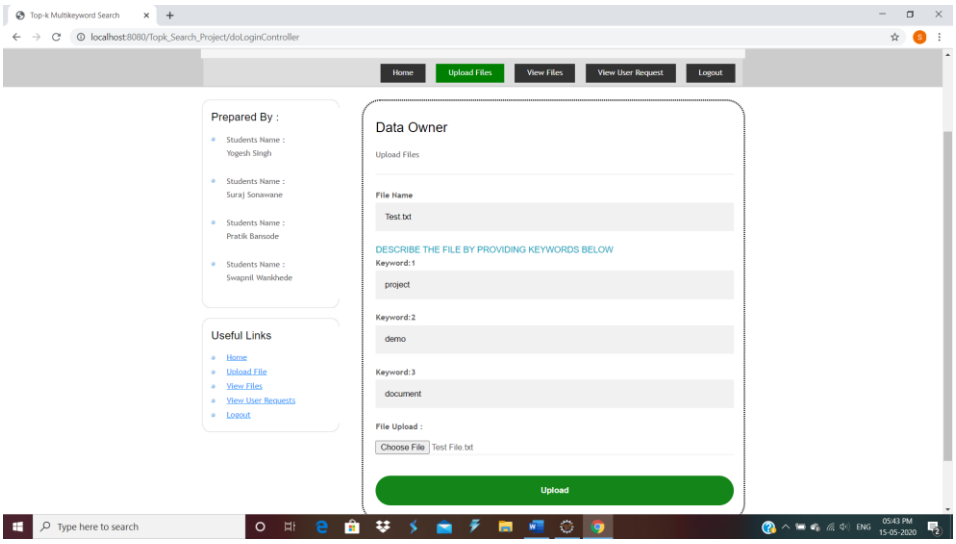


Fig.5.2 Data Owner Page.

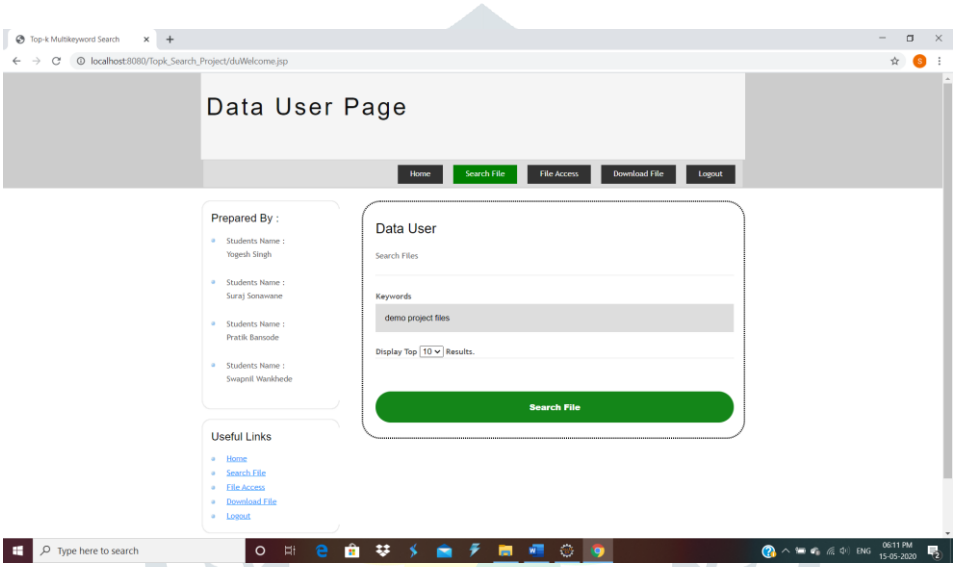


Fig.5.3 Data User page.

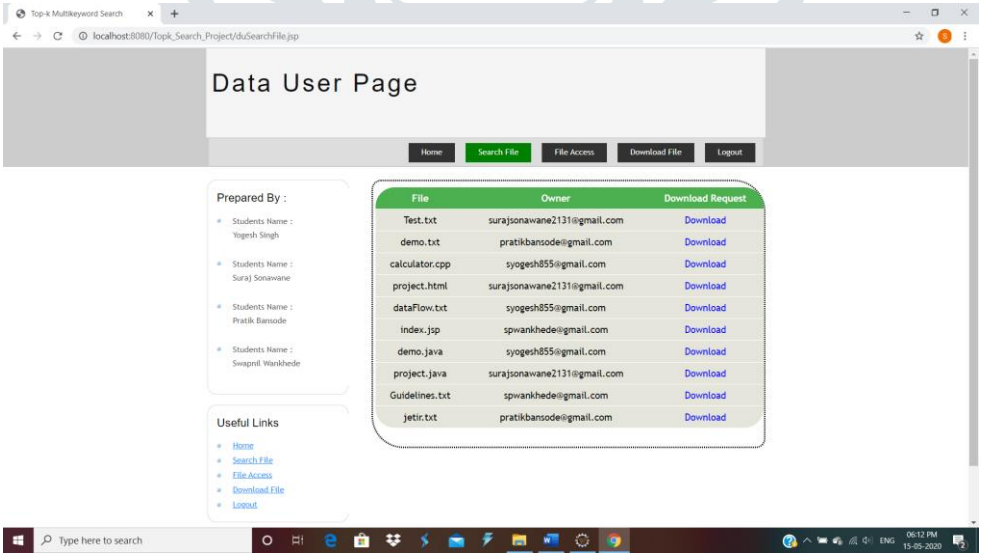


Fig.5.4 Top k search results.

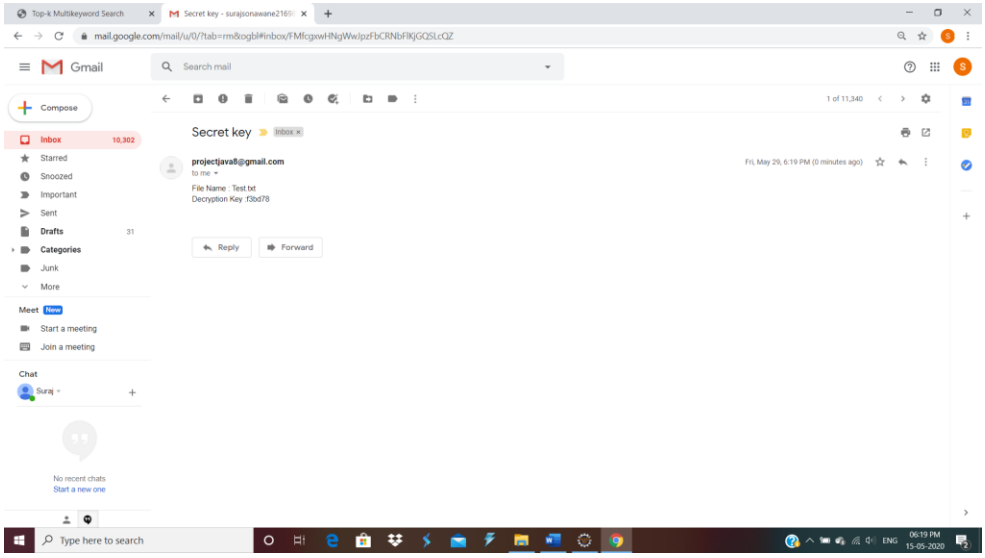


Fig.5.5 Verification Key sent on mail after Request approval.

VI. CONCLUSION

In this paper, we focused on improving the efficiency and the security of Multi-keyword top-k search over encrypted data. The user can effectively store their personal documents on cloud while preserving privacy of their documents and can retrieve them by sending a query consisting of multiple keywords. Privacy will be achieved by encrypting the queries. In response to the user's search query, system will match the keywords from query to the documents using "keyword-matching principle". Top ranked documents will get fetched based on number of keywords matched and different parameters like most downloaded, recently uploaded. Then, in order to improve the search efficiency, we design the group Multi-keyword top-k search scheme, which divides the dictionary into multiple groups and only needs to store the top-k documents. Thus, we implemented secure file sharing system which helps to securely share data.

VII. REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient con_x0002_structions," in Proceedings of the 13th ACM Conference on Computer and Communications Security.

[2] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in Cryptology_x0002_Eurocrypt 2014.

[3] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data, IEEE Transactions on Parallel and Distributed Systems.

[4] Ayad Ibrahim, Hai Jin, Ali A.Yassin, DeqingZou, "Secure Rank Ordered Search of Multi-Keyword Trapdoor over Encrypted Cloud Data" IEEE Asia-Pacific Services Computing Conference 2018.

[5] W. M. Liu, L. Wang, P. Cheng, K. Ren, S. Zhu, and M. Debbabi, "Privacy-preserving traffic padding in web-based applications."

[6] H. Li, Y. Yang, T. H. Luan, X. Liang, L. Zhou, and X. Shen, "Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data".

[7] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. Appl. Cryptography Netw. Security, Yellow Mountain.

[8] Z. Xu, W. Kang, R. Li, K. Yow, and C. Xu, "Efficient Multi keyword ranked query on encrypted data in the cloud," in Proc. IEEE 19th Int. Conf. Parallel Dist. Syst., Singapore, pp. 244–251.

[9] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings.

[10] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in INFOCOM, 2017.