

# Sensible Keyword Search for Reliable Access Management in Secure Cloud Storage

Syed Suhaila<sup>1</sup>

Assistant Professor, Department of Computer Science Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi, India<sup>1</sup>

**Abstract:** In the contemporary landscape characterized by extensive data utilization, cloud computing has emerged as a novel computing paradigm that facilitates the sharing of computing resources via the internet. The defining features of cloud computing include on-demand self-service, location-independent network access, ubiquitous network connectivity, and a pay-per-use model. These appealing attributes have led both private and public organizations to outsource substantial volumes of data to cloud storage solutions. Consequently, organizations are increasingly motivated to transition their data from local environments to centralized commercial public cloud servers. By outsourcing their data to the cloud, users can alleviate the burdens associated with storage maintenance. However, despite the numerous advantages of migrating data to cloud storage, significant security concerns arise, causing data owners to hesitate when it comes to transferring sensitive information. This situation results in a shift of data control towards cloud service providers. To address these security challenges, data owners are compelled to encrypt their data prior to outsourcing. While encryption enhances data security, it simultaneously diminishes data efficiency, as searching through encrypted data proves to be complex. Traditional search techniques applicable to plaintext data cannot be utilized for encrypted data. Current solutions primarily support exact keyword searches, lacking support for semantic searches. In this project, we propose a semantic multi-keyword ranked search system that incorporates verifiable outsourced decryption. To enhance search efficiency, this system integrates semantic search capabilities through the implementation of fuzzy search techniques.

**Keywords:** Encryption, sensitive, multi-keyword, service providers.

## I. INTRODUCTION

With the increasing prevalence of Cloud Computing, there is a growing centralization of sensitive information within the cloud. To safeguard data privacy, it is essential to encrypt sensitive data prior to outsourcing, which complicates effective data utilization. While traditional searchable encryption methods enable users to securely search encrypted data using keywords, these methods are limited to Boolean searches and do not account for the relevance of data files. This limitation presents two significant challenges in the context of Cloud Computing. Firstly, users, who may lack prior knowledge of the encrypted cloud data, must manually process each retrieved file to identify those that align with their interests. Secondly, the necessity to retrieve all files containing the queried keyword leads to excessive network traffic, which is particularly undesirable in the current pay-as-you-use cloud model. Our objective is to develop an efficient system that allows any authorized user to conduct searches on a remote database using multiple keywords, without disclosing either the keywords being searched or the contents of the retrieved documents. Unlike previous systems that assume only the data owner queries the database, our proposal enables a group of users to query the database, provided they possess trapdoors for the search terms that authorize their inclusion in the queries. Furthermore, our system is designed to execute multiple keyword searches within a single query and to rank the results, allowing users to access only the most relevant matches [1-2].

The advent of cloud technology has enabled users to utilize cloud computing for accessing data from any location globally without temporal limitations. However, the data stored in the cloud may include sensitive information that requires protection. To ensure privacy, data holders must encrypt their information before uploading it to the cloud. Once encrypted, the data is secured; however, challenges arise for users. The user-friendly nature of search engines like Google allows individuals to input queries in plain text. Additionally, there is ongoing development of a technique that facilitates simultaneous searches for multiple keywords. Effective searching is enhanced through the application of logical operations such as AND, OR, and NOT. Moreover, there is a focus on fraud detection, which involves alerting data owners when fraudulent activities are detected. The vision of presenting computing as a utility is being realized through cloud computing. To benefit from improved operational quality, cloud users can store their data remotely in the cloud. The objective of this research is to determine the relevance score of keywords, thereby improving the accuracy of keyword searches. Additionally, the study will introduce a robust searching framework that accommodates multiple-

keyword searches, allowing for complex queries that incorporate logical operations. Furthermore, the research aims to establish an advanced indexing system by organizing sub-dictionaries effectively. Lastly, the study seeks to enhance system reliability by implementing a notification mechanism that alerts data owners when unauthorized users attempt to modify their data in the cloud [3].

The current system is inadequate in delivering a scalable search capability that enables the retrieval of multiple keywords from encrypted cloud data. The proposed research initiative addresses these challenges by allowing users to conduct searches for multiple keywords within encrypted cloud data. Furthermore, the system will incorporate fault tolerance features. The analytical model encompasses a thorough examination of its components, which can be categorized into analytical analysis and experimental analysis. Analysis involves breaking down complex structures into simpler components for enhanced comprehension. Analytical analysis is subdivided into mathematical modelling, while implementation falls under experimental analysis. The issue at hand pertains to privacy-preserving keyword search within a private database model, where documents are encrypted using secret keys that remain unknown to the actual database holder (i.e., the Cloud Server). The data owner, who generates and collects the information within the database, may lack the resources or willingness to manage the database. Users are individuals within a group authorized to access certain portions of the database's information. The server, typically a professional entity such as a cloud service, provides information services to these authorized users. It is often essential for the server to remain unaware of the database's content, the search terms used in queries, and the documents retrieved [4-5].

## **II. RELATED WORK**

Liang et al. conducted a study on cloud resource allocation within a multi-domain mobile cloud system characterized by several key features: 1) the arrivals and departures of mobile cloud services adhere to a Poisson distribution; 2) the cloud's available resources fluctuate over time; and 3) current resource allocation decisions can significantly influence future choices. In a multi-domain cloud environment, neglecting the interplay between present and future resource allocation decisions can lead to a decline in overall system performance. To develop a robust resource allocation model for geo-based mobile cloud computing, we propose a decision support system that takes into account cloud system resources, profit generation, and the degree of engagement (DoE) of mobile users. The primary aim of this paper is to enhance the overall rewards for both the cloud system and its mobile users. In our proposed model, the arrivals and departures of mobile application services are treated as random events, resulting in changes to the cloud's resource state. In line with the semi-Markov decision process (SMDP) framework, the decision epoch can be established at the occurrence of any random event. Consequently, we begin by evaluating the system rewards within a cloud domain while considering inter-domain resource transfers through an SMDP model. The resource allocation decision model we present seeks to optimize resource distribution across mobile cloud service domains. Our findings indicate that this solution not only enhances the utilization of cloud system resources but also improves the degree of engagement for mobile users [6].

Mahmoud et al. identified a hotspot phenomenon that leads to a significant inconsistency in network traffic patterns, primarily due to a high volume of packets emanating from a limited geographical area. Hotspots may arise for various reasons, such as when pandas congregate in high densities or remain in a specific location due to the presence of food, water, shade, or shelter. Furthermore, we establish a realistic adversary model that assumes the adversary possesses a partial view of the network traffic by deploying a series of monitoring devices at various observation points. Each device gathers traffic data, which includes the content of packets, the coordinates of the sending node, and the timestamps of packet transmission. Utilizing this model, we introduce an innovative attack termed Hotspot-Locating, wherein the adversary exploits the traffic inconsistencies generated by hotspots to pinpoint the locations of pandas by analyzing the data collected from the observation points through traffic analysis methods, such as examining packet sending rates and correlations. Lastly, we propose a cloud-based strategy designed to effectively safeguard the location privacy of source nodes against the Hotspot-Locating attack. This is achieved by generating a cloud with an irregular shape of deceptive traffic, which mitigates the inconsistencies in traffic patterns caused by hotspots and conceals the actual source node among the nodes that constitute the cloud. The fabricated packets also facilitate the real source node in transmitting sensed data anonymously to a fictitious source node selected from the cloud's nodes for delivery to the Sink flexible delegation, permitting any subset of the cipher texts to be decrypted using a fixed-size decryption key [7].

Shen et al. explored the use of geo-distributed clouds in e-health monitoring systems, identifying the minimization of service delay as a significant research challenge due to user mobility and the decentralized management of cloud servers. Some studies recommend leveraging the location data of cloud servers for resource allocation to mitigate service delays; however, they often overlook the specific service delay requirements inherent to e-health monitoring systems. Developing a tailored distributed resource allocation strategy for such systems holds considerable research

importance and practical implications. Additionally, privacy preservation emerges as a critical concern for users utilizing cloud computing in e-health monitoring, as the transmission of health data to remote servers may expose it to various security threats over long-distance and insecure channels. This paper presents an e-health monitoring system that is underpinned by geo-distributed clouds, which comprise numerous cloud servers strategically located across a vast area. The proposed system is divided into two components: a resource allocation framework for the servers and a traffic-shaping algorithm designed for users [8].

Cao has established a multi-keyword ranked search system for encrypted cloud data (MRSE) that maintains stringent privacy standards within the cloud computing framework. Among the various multi-keyword semantics available, we have selected the effective similarity measure known as "coordinate matching," which aims to maximize the number of matches to assess the relevance of data documents in relation to the search query. Specifically, we employ "inner product similarity," which counts the number of query keywords present in a document, to quantitatively assess the similarity of that document to the search query. In the process of index construction, each document is linked to a binary vector serving as a subindex, where each bit indicates the presence of the corresponding keyword within the document. The search query is similarly represented as a binary vector, with each bit denoting whether the corresponding keyword is included in the search request. Consequently, the similarity can be precisely determined by calculating the inner product of the query vector and the data vector. However, outsourcing either the data vector or the query vector directly would compromise index privacy or search privacy. To address the challenge of facilitating such multi-keyword semantics without compromising privacy, we propose a foundational concept for the MRSE that utilizes secure inner product computation, adapted from a secure k-nearest neighbor (KNN) methodology. We then present two significantly enhanced MRSE schemes in a systematic manner to fulfil various stringent privacy requirements across two threat models with elevated attack capabilities. [9].

### III. PROPOSED METHOD

The proposed system architecture is shown in Figure 1. It has five modules: Cloud framework construction, File upload, Index table construction, Search results and Verifiable outsourced decryption.

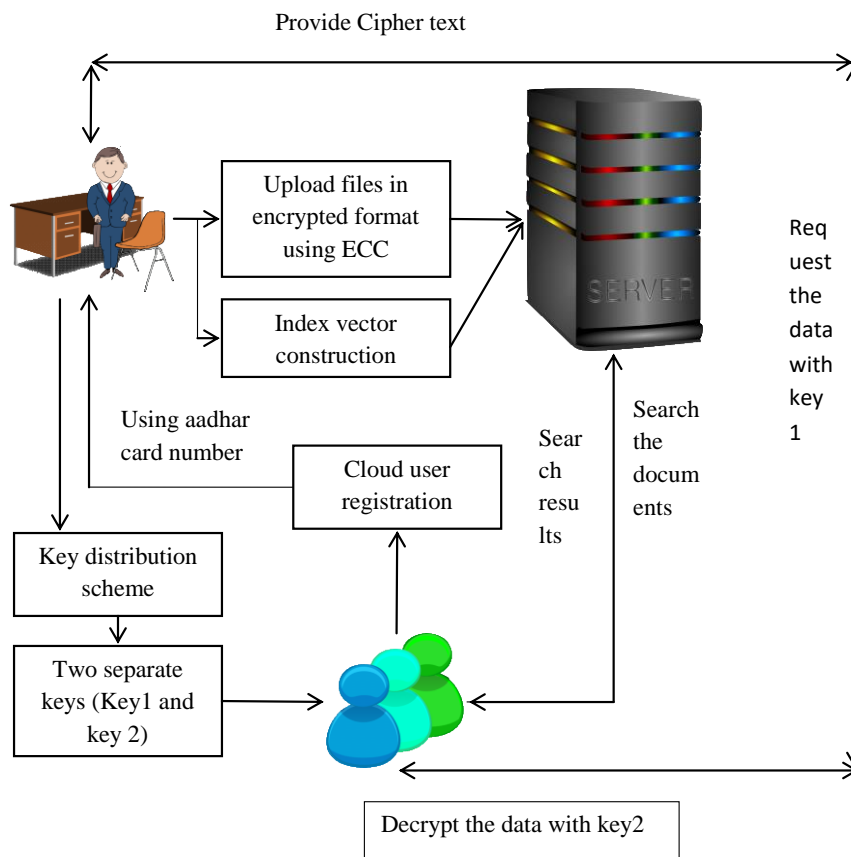


Fig 1: Proposed System Architecture

*Cloud framework construction:*

The concept of computing as a utility has long been envisioned, allowing cloud customers to remotely store their data in the cloud and access high-quality applications and services on demand from a shared pool of configurable computing resources. This model offers significant flexibility and cost savings, prompting both individuals and organizations to transition their complex local data management systems to the cloud. This document addresses the challenge of conducting multi-keyword ranked searches over encrypted cloud data for the first time and sets forth a comprehensive set of privacy requirements for a secure cloud data utilization system. The module comprises three user types: the cloud owner, the cloud server, and the end users. It facilitates the registration of the owner's details, including login information, and authorizes users to access the system. To enhance the security of user data, login credentials are encrypted and subsequently decrypted by the server to prevent eavesdropping. The server is responsible for storing files in cloud storage, while users can search for files using specific keywords.

*File upload:*

This module enables the owner to securely upload files through the use of the Elliptic Curve Cryptography (ECC) algorithm. This mechanism safeguards the files against unauthorized access. Upon logging into the system, the data owner can incorporate information obtained from web crawling, with the data organized in a manner that facilitates easy access. Given the potentially large volume of data, it is essential to maintain an appropriate structure for storage. ECC is a form of public key cryptography, where each participant in the communication possesses a pair of keys: a public key and a private key, along with a series of operations linked to these keys for executing cryptographic functions. The private key remains confidential to the specific user, while the public key is shared among all participants in the communication. Certain public key algorithms may necessitate a set of predefined constants to be known by all devices involved in the communication.

*Index table construction:*

An index is established as a compilation of mappings corresponding to each keyword. The compilation for a specific keyword includes the following details:

1. Identifiers of the files that contain the specific keyword
2. The term frequency for each file, indicating how many times the keyword appears within that file, which reflects the keyword's significance in that context.
3. The length of each file
4. A relevance score for each file
5. The total number of files that include the specific keyword. Data structures such as tables may be employed to organize this information. The term frequency, file length, and the number of files associated with the keyword are utilized to compute the relevance score for each file through scoring mechanisms that will be elaborated upon in the Ranking modules. When a data file is saved, it undergoes pre-processing to create an index that encompasses the aforementioned details, derived from the keywords extracted using various string matching algorithms.

The procedure for index creation is outlined as follows:

1. For each  $w_i$  within the keyword set  $W$ , generate  $F(w_i)$ , which represents the file identifiers that include  $w_i$ .
2. For each  $w_i$  in  $W$ , for  $1 \leq j \leq |F(w_i)|$ 
  - 2.1. Compute the score of the file  $F_{ij}$  (utilizing the scoring mechanisms to be discussed later) and record it as
  - 2.2. Store it alongside the file identifier  $id(F_{ij})$  and the length of the file  $|F_{ij}|$  in the index list  $I(w_i)$  for the specific keyword  $w_i$
  - 2.3. Update the total count of files containing the keyword in the index list as  $(I(w_i)||N)$ .

*Search results:*

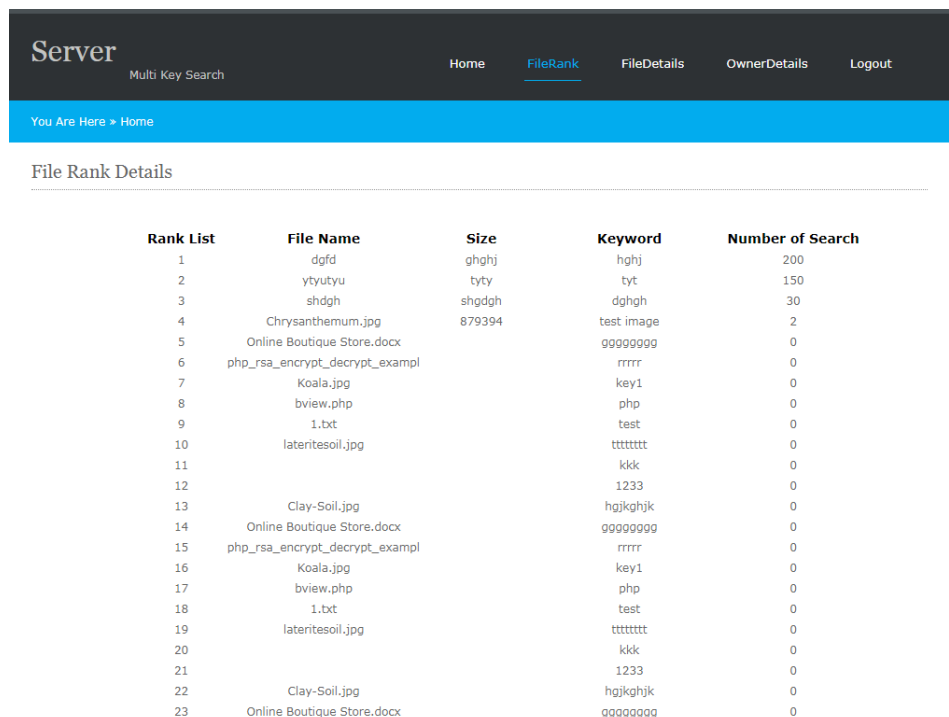
This module facilitates users in locating frequently searched files through a ranking search mechanism. Once documents are stored and indexed, the subsequent critical task is to rank them based on available details, enabling users to retrieve the top "k" most pertinent documents. To achieve this, it is necessary to compute a numerical score for each file. Within the information retrieval community, the predominant ranking functions are derived from the TF-IDF principle, where TF denotes Term Frequency, indicating the frequency of a keyword's occurrence within a file, and IDF signifies Inverse Document Frequency, which is calculated as the ratio of the number of files containing the keyword to the total number of files on the server. When a user requests data, ranking is performed on the requested information utilizing a fuzzy search algorithm. The ranking process employs the coordinate matching principle. Following the ranking, users receive the anticipated results for their queries.

*Verifiable outsourced decryption:*

In order to obtain the file, the client must submit a request to the owner. Should the owner approve this request, the client will gain access to the file. The data stored on the server is organized in a manner that facilitates easy querying, and this approach also offers instant search assistance, enabling users to swiftly locate the information they need. During the access procedure, the user seeking file access is required to provide attributes in accordance with the access policy established by the owner. If the user's attributes align with the access policy, meaning the policy is satisfied, the system initiates the first decryption process using the access policy, facilitated by a proxy server. Additionally, the checksum is decrypted at this stage. The partially decrypted file remains with the proxy server, which then requests the key from the user. If the provided key is accurate, the second decryption occurs, resulting in the fully decrypted file and key being transmitted to the user.

## IV. RESULTS AND DISCUSSION

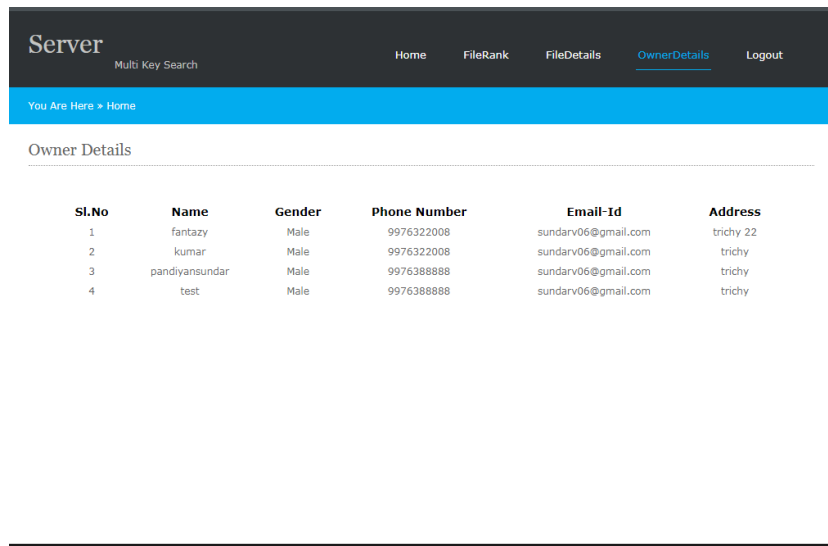
The results of the detailed investigations are discussed in this section. The results are shown in Figures 2-5. In the realm of cloud computing, storage and computational resources are provided as scalable and elastic services via the Internet. By outsourcing data services to the cloud, organizations can benefit from cost reductions and a more streamlined local IT management, as the physical hosting and maintenance of cloud infrastructures are handled by the service providers. To mitigate the risk of data breaches by these providers, data owners often choose to encrypt sensitive information, such as health records and financial transactions, prior to cloud outsourcing, while retaining the decryption keys for themselves and authorized personnel. This approach, however, complicates data utilization. For instance, searching for relevant documents within an encrypted dataset stored in the cloud may necessitate downloading and decrypting the entire dataset, which is impractical for large volumes of data [10-11]. Consequently, there is significant interest in developing mechanisms that enable users to perform searches directly on encrypted data in the cloud computing landscape. Despite advancements, efficient multi-keyword fuzzy search over encrypted data continues to pose challenges. It is important to note that efforts to facilitate search over encrypted data encompass not only information retrieval methodologies, such as sophisticated data structures for searchable indices and efficient search algorithms, but also the careful design of cryptographic protocols to safeguard the security and privacy of the entire system. While multi-keyword search and fuzzy search have been executed independently, their integration has not resulted in a secure and efficient multi-keyword fuzzy search framework. In this paper, we introduce a novel concept for achieving multi-keyword fuzzy search (specifically conjunctive keywords). Unlike existing multi-keyword search frameworks, our approach does not require a predefined keyword dictionary. The inherent fuzziness of the keywords is effectively captured through an innovative data structure and algorithmic design without expansion.



The screenshot shows a web application titled 'Server' with a navigation menu including 'Home', 'FileRank', 'FileDetails', 'OwnerDetails', and 'Logout'. Below the navigation is a breadcrumb trail 'You Are Here > Home'. The main content area is titled 'File Rank Details' and contains a table with the following data:

Rank List	File Name	Size	Keyword	Number of Search
1	dgfd	ghghj	ghghj	200
2	ytyutyu	tyty	tyt	150
3	shdgh	shgdgh	dghgh	30
4	Chrysanthemum.jpg	879394	test image	2
5	Online Boutique Store.docx		gggggggg	0
6	php_rsa_encrypt_decrypt_exampl		rrrrr	0
7	Koala.jpg		key1	0
8	bview.php		php	0
9	1.txt		test	0
10	lateritesoil.jpg		ttttttt	0
11			kkk	0
12			1233	0
13	Clay-Soil.jpg		hgjkgjhk	0
14	Online Boutique Store.docx		gggggggg	0
15	php_rsa_encrypt_decrypt_exampl		rrrrr	0
16	Koala.jpg		key1	0
17	bview.php		php	0
18	1.txt		test	0
19	lateritesoil.jpg		ttttttt	0
20			kkk	0
21			1233	0
22	Clay-Soil.jpg		hgjkgjhk	0
23	Online Boutique Store.docx		gggggggg	0

Fig 2: File Rank Details



**Server**  
Multi Key Search

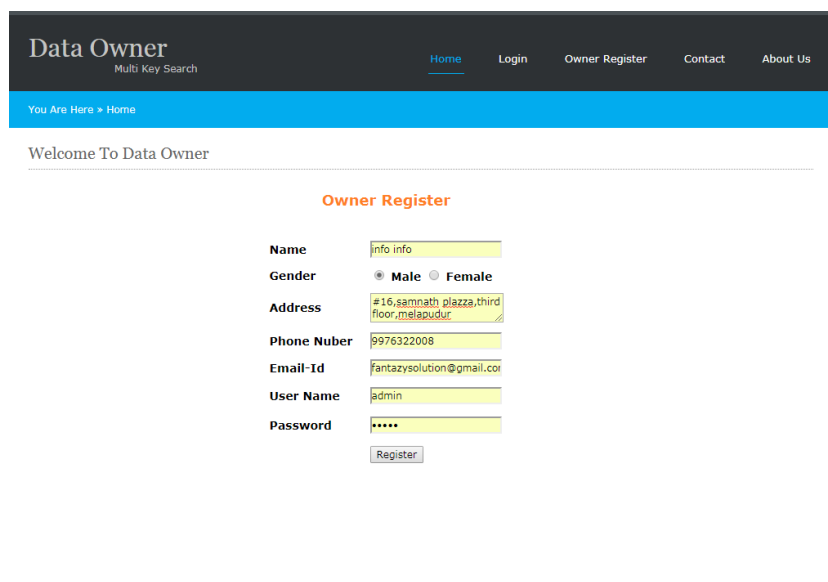
Home FileRank FileDetails OwnerDetails Logout

You Are Here > Home

Owner Details

Sl.No	Name	Gender	Phone Number	Email-Id	Address
1	fantazy	Male	9976322008	sundarv06@gmail.com	trichy 22
2	kumar	Male	9976322008	sundarv06@gmail.com	trichy
3	pandiyansundar	Male	9976388888	sundarv06@gmail.com	trichy
4	test	Male	9976388888	sundarv06@gmail.com	trichy

Fig 3: Owner Details



**Data Owner**  
Multi Key Search

Home Login Owner Register Contact About Us

You Are Here > Home

Welcome To Data Owner

**Owner Register**

Name: info info

Gender:  Male  Female

Address: #16,samnath plaza,third floor,melapudur

Phone Nuber: 9976322008

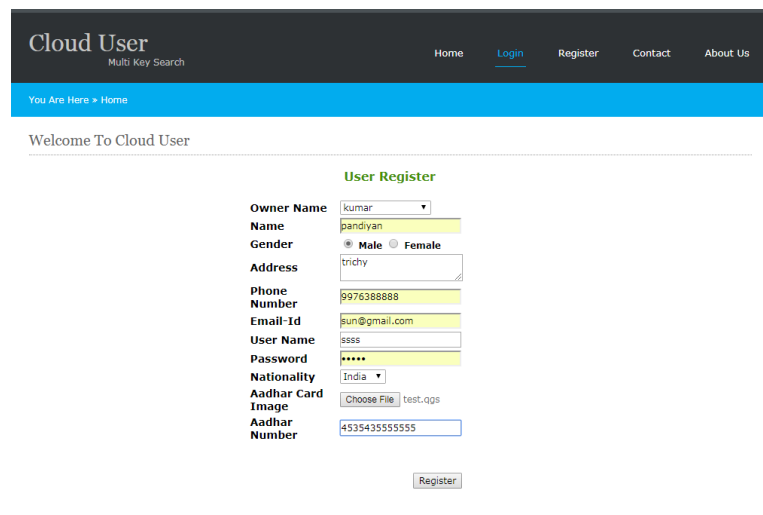
Email-Id: fantazysolution@gmail.com

User Name: admin

Password: \*\*\*\*\*

Register

Fig 4: Owner Registration



**Cloud User**  
Multi Key Search

Home Login Register Contact About Us

You Are Here > Home

Welcome To Cloud User

**User Register**

Owner Name: kumar

Name: pandiyan

Gender:  Male  Female

Address: trichy

Phone Number: 9976388888

Email-Id: sun@gmail.com

User Name: ssss

Password: \*\*\*\*\*

Nationality: India

Aadhar Card Image: Choose File test.qgs

Aadhar Number: 4535435555555

Register

Fig 5: Owner Registration

**V. CONCLUSION**

This article addresses the complex issue of multi-keyword fuzzy search within encrypted data. We have introduced and integrated several innovative designs aimed at simultaneously resolving the challenges of multiple keyword searches and fuzzy searches with high efficiency. Our novel approach utilizes functions to construct the file index, offering an effective solution for secure fuzzy searches involving multiple keywords. Furthermore, the fuzzy search mechanism is employed to assess the similarity among keywords, while secure inner product computation is utilized to derive similarity scores for result ranking. We have developed both a fundamental scheme and an enhanced scheme to accommodate varying security requirements, along with a semantic scheme to extract pertinent results from encrypted cloud data. Additionally, we have proposed a concrete Attribute-Based Encryption (ABE) scheme featuring verifiable outsourced decryption, designed for an IT firm to securely store and access documents, demonstrating its security and verifiability. This scheme has shown to be more efficient than traditional ABE and ABE with outsourced decryption across all evaluated aspects. The intermediate ciphertext can be converted into plaintext by a proxy server, incurring minimal computational overhead. The security of an ABE system with outsourced decryption ensures that adversaries, including malicious proxies, cannot glean any information about the encrypted message; however, it does not guarantee the accuracy of the transformation performed by the cloud. We also introduce a new requirement for ABE with outsourced decryption: verifiability. Comprehensive theoretical security analysis and experimental evaluations using real-world datasets have been conducted to validate the practicality of our proposed scheme. The ranking method we proposed has proven efficient in returning highly relevant documents corresponding to the submitted search terms. We have implemented the entire scheme, and extensive experimental results from this implementation demonstrate the effectiveness and efficiency of our solution.

**REFERENCES**

- [1]. T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacy-preserving data aggregation without secure channel: Multivariate polynomial evaluation," in Proc. IEEE INFOCOM, 2012, pp. 2634–2642.
- [2]. H. Li, Y. Dai, L. Tian, and H. Yang, "Identity-based authentication for cloud computing," in Cloud Computing, Springer, 2010, pp. 157–166.
- [3]. R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multi-keyword ranked query over encrypted data in cloud computing," Future Generation Comput.Syst., vol. 30, pp. 179–190, 2008.
- [4]. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. S&P, IEEE, 2009, pp. 44–55.
- [5]. Y. H. Hwang and P. J. Lee, Public Key Encryption with Conjunctive Keyword Search and Its Extension to a Multi-user System. In Proceedings of International Conference on Pairing-Based Cryptography, (2007).
- [6]. H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdp based service model for interdomain resource allocation in mobile cloud networks," IEEE Trans. Veh. Technol., vol. 61, no. 5, pp. 2222–2232, Jun. 2014.
- [7]. M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 10, pp. 1805–1818, Oct. 2013.
- [8]. Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting geo-distributed clouds for e-health monitoring system with minimum service delay and privacy preservation," IEEE J. Biomed. Health Inform., vol. 18, no. 2, pp. 430–439, Mar. 2012.
- [9]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2011.
- [10]. D. Boneh and B. Waters, Conjunctive, subset, and range queries on encrypted data, In Proceedings of the 4th Theory of Cryptography Conference, (2007).
- [11]. N. Cao, C. Wang, M. Li, K. Ren and W. Lou, Privacy-preserving multi-keyword ranked search over encrypted cloud data, In Proceedings of IEEE INFOCOM, (2011)