Novateur Publications
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
Volume 11, Issue 8, August - 2024

PRIVACY-PRESERVING MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA

Miss. Neha Shrirang Lad
Miss. Deepali Deelip Suryawanshi
Adarsh Institute of Technology and Research Centre, Vita

Urmila Ravindra Patil
Assistant Professor, Adarsh Institute of Technology and Research Centre, Vita

Article History: Received on: 05/06/2024
Accepted on: 14/08/2024

DOI: https://doi.org/10.26662/ijiert.v11i8.pp1-12

ABSTRACT

The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.

INTRODUCTION

We are living in a highly networked environment, where huge amounts of data are stored in remote, but not necessarily trusted servers. There are several privacy issues regarding to accessing data on such servers; two of them can easily be identifed: sensitivity of i) keywords sent in queries and ii) the data retrieved; both need to be hidden. A related protocol, Private Information Retrieval (PIR) enables the user to access public or private databases without revealing which data he is extracting. Since privacy is of a great concern, PIR protocols have been extensively studied in the past.

In today's information technology landscape, customers that need high storage and computation power tend to outsource their data and services to clouds. Clouds enable customers to remotely store and access their data by lowering the cost of hardware ownership while providing robust and fast services. The importance and necessity of privacy preserving search techniques are even more pronounced in the cloud applications. Due to the fact that large companies that operate the public clouds like Google or Amazon may access the sensitive data and search patterns, hiding the query and the retrieved data has great importance in ensuring the privacy and security of those using cloud services.

We aim to achieve an effcient system where any authorized user can perform a search on a remote database with multiple keywords, without revealing neither the keywords he searches for, nor the contents of the documents he retrieves. Our proposed system differs from the previous works which assume that only the data owner queries the database. In contrast to previous works, our proposal facilitate that a group of users can query the database provided that they possess trapdoors for search terms that authorize the users to include them in their queries. Moreover, our proposed system is able to perform multiple keyword search in a single query and ranks the results so the user can retrieve only the top matches.

The contributions of this paper can be summarized as follows. Firstly, we provide formal definitions for the security and privacy requirements of keyword search on encrypted cloud data. Secondly, we propose an efficient ranked multi-keyword search scheme and formally prove that it is secure in accordance with the defined requirements. Thirdly, we propose a ranking method that proves to be efficient to implement and effective in returning documents highly relevant to submitted search terms. Lastly, we implement the proposed scheme and demonstrate that it is much more efficient than existing methods in literature.

RELATED WORK

The problem of Private Information Retrieval was first introduced by Chor et al. Recently Groth et al propose a multi query PIR method with constant communication rate. However, any PIR-based technique requires highly costly cryptographic operations in order to hide the access pattern. This is inefficient in the large scale cloud system and as an alternative approach, privacy preserving search is employed which aims to hide the content of the retrieved data instead of which data is retrieved.

Ogata and Kurosawa show privacy preserving keyword search protocol based on RSA blind signatures. The scheme requires a public key operation per item in the database for every query and this operation must be performed on the user side.

Freedman et al, proposed an alternative implementation for private keyword search that uses homomorphic encryption and oblivious polynomial evaluation methods. The computation and communication costs of this method are quite large since every search term in a query requires several homomorphic encryption operations both on the server and the user side.

A recent work proposed by Wang et al allows ranked search over an encrypted database by using inner product similarity. However, this work is only limited to single keyword search queries.

One of the closest methods to our solution is proposed by Cao et al. Similar to our approach presented here, it proposes a method that allows multi-keyword ranked search over encrypted database. In this method, the data owner needs to distribute a symmetric key which is used in trapdoor generation to all authorized users. Additionally, this work requires keyword fields in the index. This means that the user must know a list of all valid keywords and their positions as a compulsory information to generate a query. This assumption may not be applicable in several cases. Moreover, it is not efficient due to matrix multiplication operations of square matrices where the number of rows are in the order of several thousands.

Wang et al propose a trapdoorless private keyword search scheme, where their model requires a trusted third party which they named as the Group Manager. We adapt their indexing method to our scheme, but we use a totally different encryption methodology to increase the security and efficiency of the scheme.

EXISTING SYSTEM

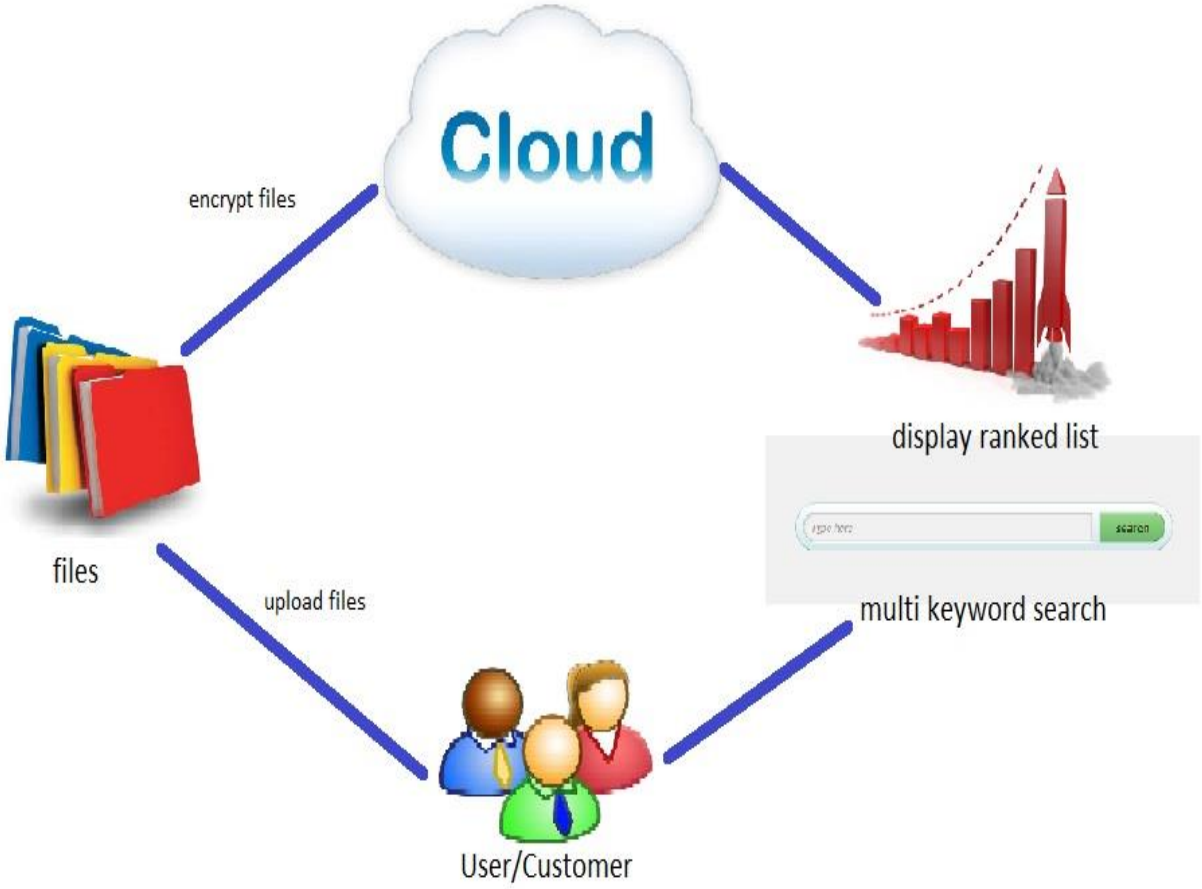
Today, there are large number of data users and documents in cloud. It is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results.

Disadvantage:

- » Single-keyword search without ranking
- » Boolean- keyword search without ranking
- » Single-keyword search with ranking

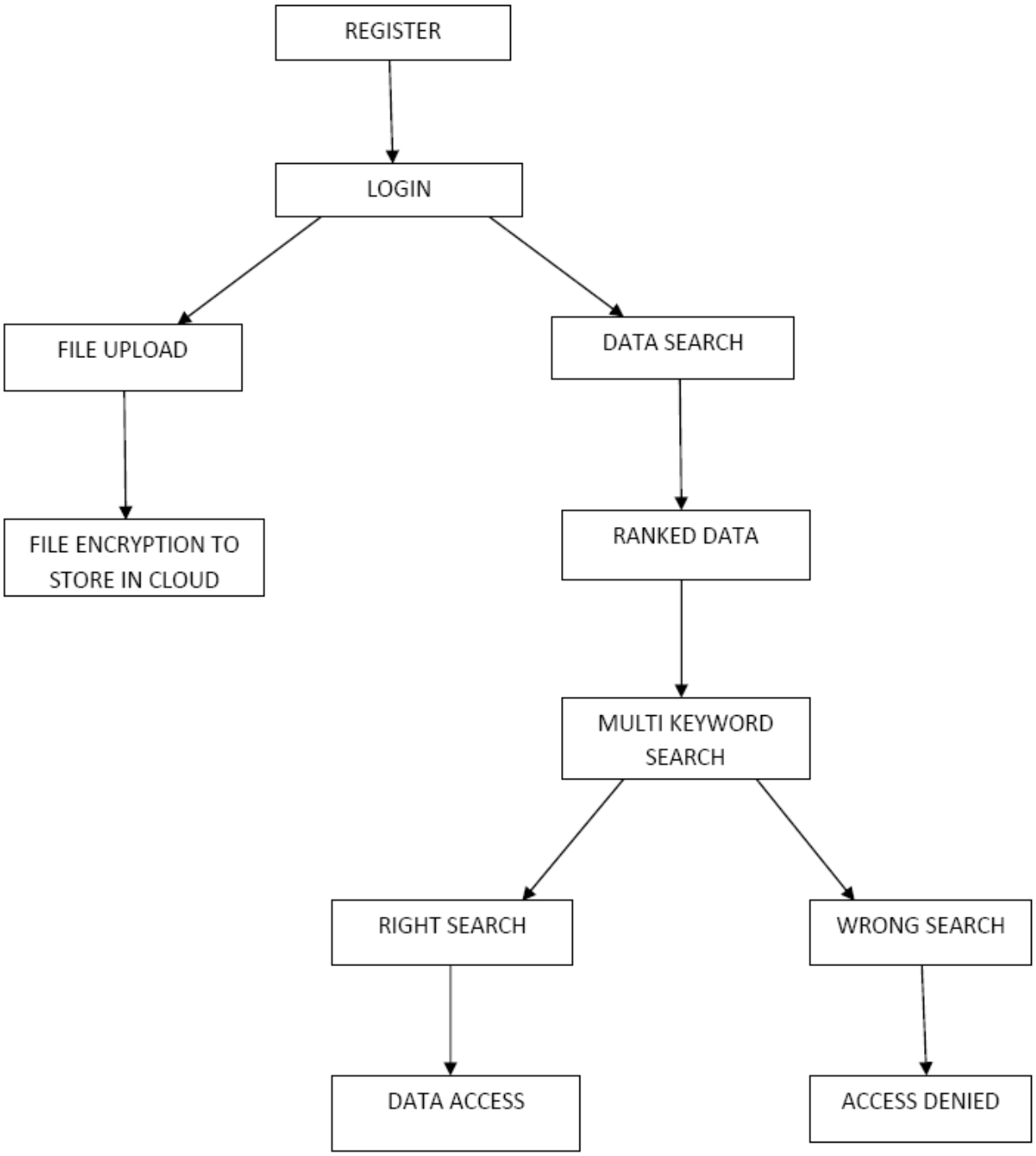
PROPOSED SYSTEM

We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching".



Novateur Publications
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
Volume 11, Issue 8, August - 2024

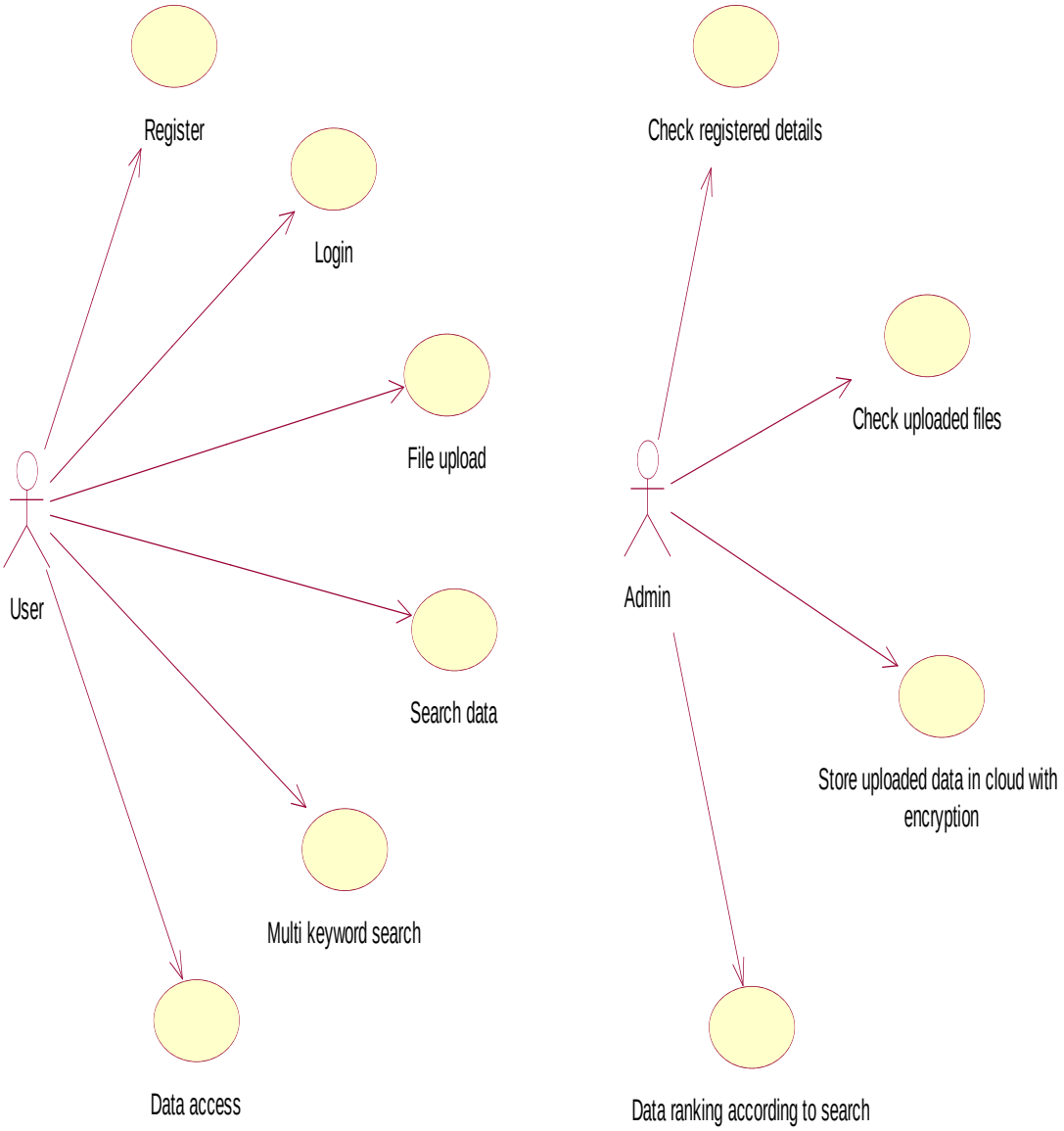
SYSTEM FLOWCHART



Novateur Publications
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
Volume 11, Issue 8, August - 2024

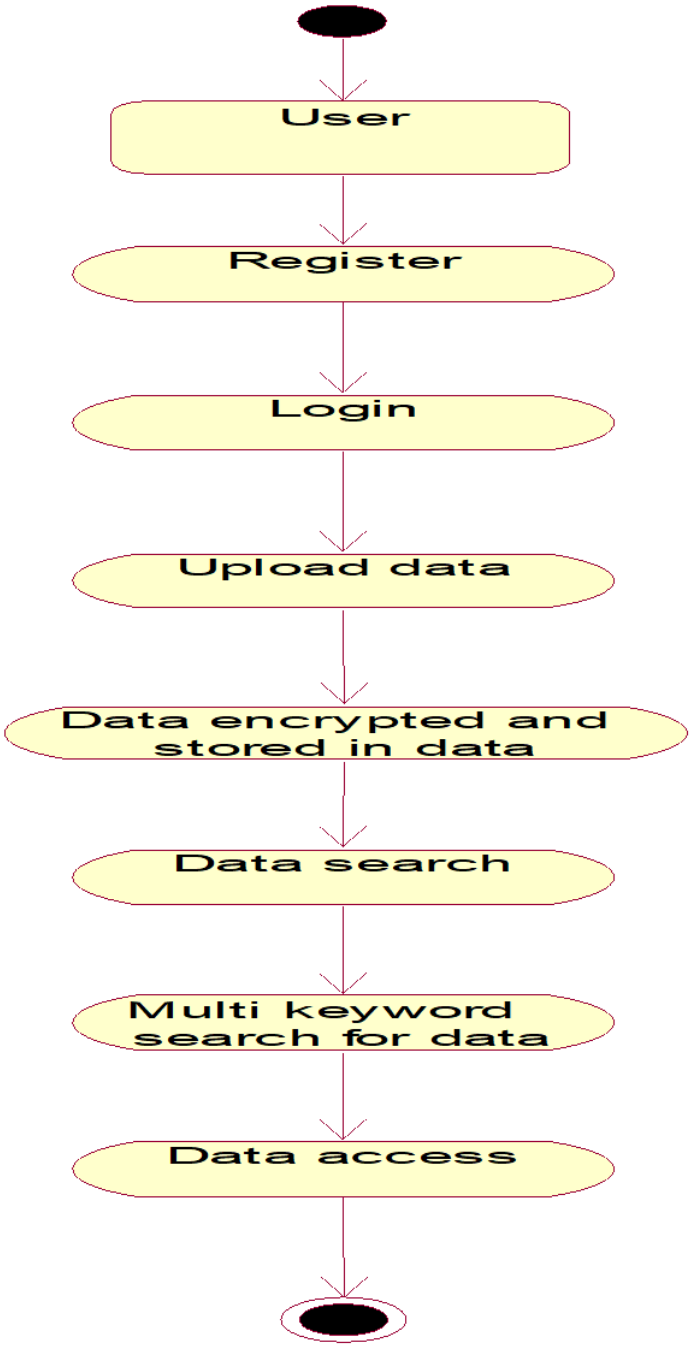
UML DIAGRAMS

USE CASE

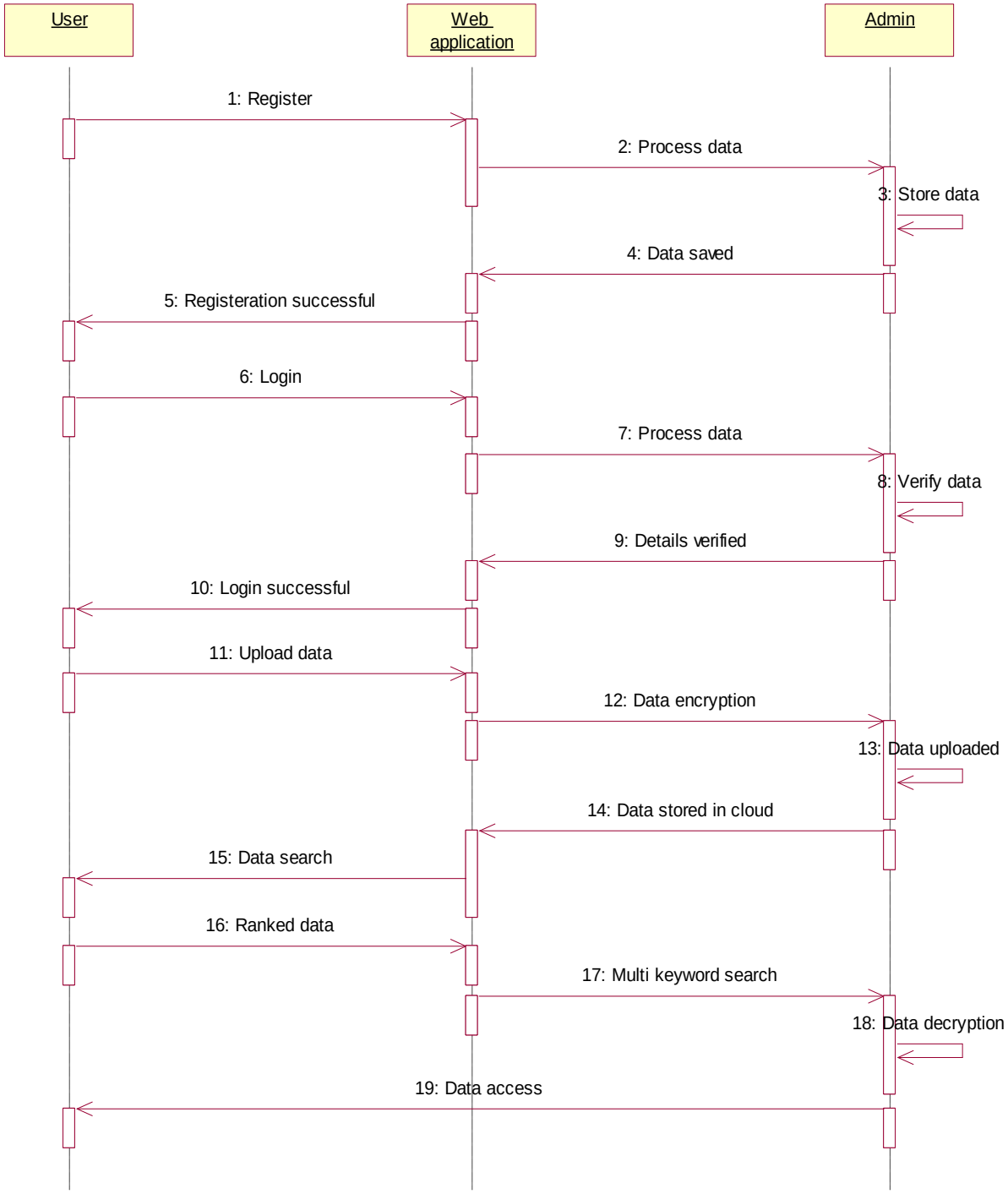


Novateur Publications
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
Volume 11, Issue 8, August - 2024

ACTIVITY

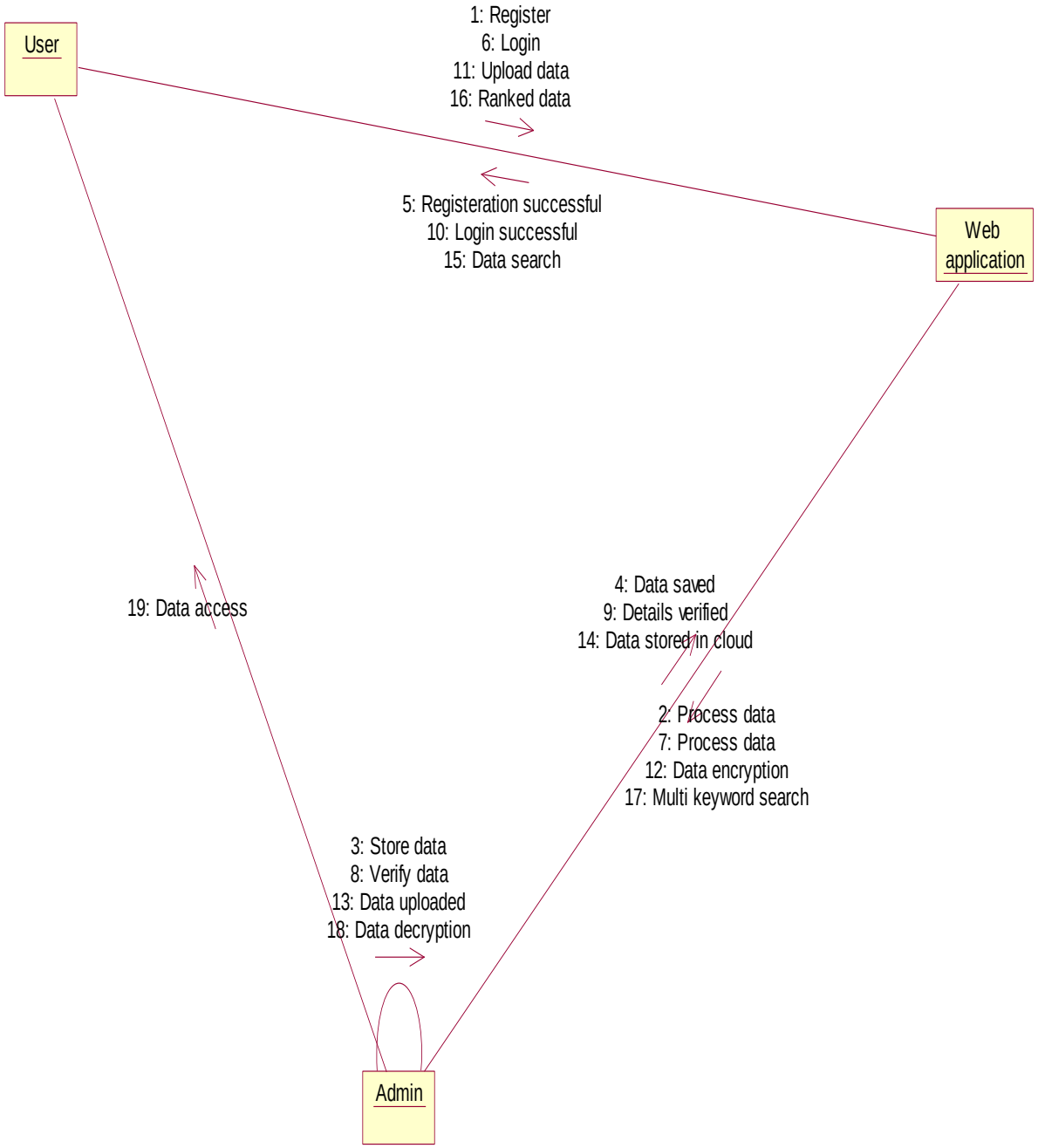


SEQUENCE

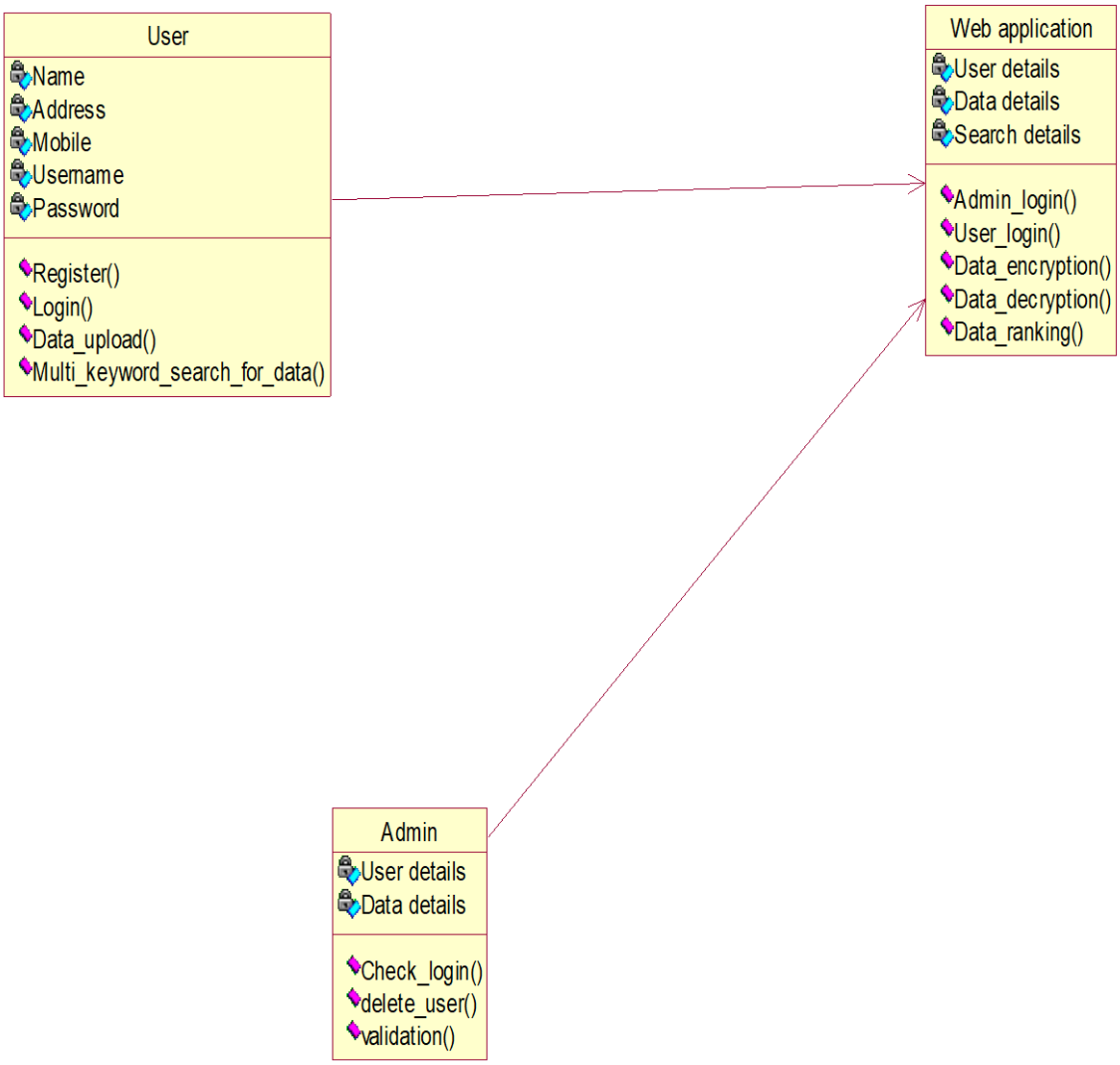

User
Web application
Admin
1: Register
2: Process data
3: Store data
4: Data saved
5: Registeration successful
6: Login
7: Process data
8: Verify data
9: Details verified
10: Login successful
11: Upload data
12: Data encryption
13: Data uploaded
14: Data stored in cloud
15: Data search
16: Ranked data
17: Multi keyword search
18: Data decryption
19: Data access

Novateur Publications
International Journal of Innovations in Engineering Research and Technology [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
Volume 11, Issue 8, August - 2024

COLLABORATION



CLASS



SYSTEM REQUIREMENTS

Hardware Requirements:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Floppy Drive : 1.44 Mb.
- Monitor : 14' Colour Monitor.
- Mouse : Optical Mouse.
- Ram : 512 Mb.
- Keyboard : 101 Keyboard.

Software Requirements:

- Operating system : Windows XP.
- Coding Language : ASP.Net with C#
- Data Base : SQL Server 2005.

METHODOLOGY

To enable ranked search for effective utilization of outsourced cloud data under the afore mentioned model, our system design should simultaneously achieve security and performance guarantees as follows.

- Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.
- Privacy-Preserving: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements specified in section III-B.
- Efficiency: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

Search index files for the documents in the database are generated by the data owner using secret keys. A user who wants to include a search term in his query, needs the corresponding trapdoor from the data owner since he does not know the secret keys used in the index generation. Asking for the trapdoor openly would violate the privacy of the user against the data owner, therefore a technique is needed to
hide the trapdoor asked by the user from the data owner.
There are various modules used in the system for the multi keyword ranked search.

1. Encrypt Module
2. Client Module
3. Multi-keyword Module
4. Admin Module

Encrypt Module:

This module is used to help the server to encrypt the document using Symmetric encrytion Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

Client Module:

This module is used to help the client to search the file using the multiple key words concept and get the accurate result list based on the user query. The user is going to select the required file and register the user details and get activation code in mail from the “customerservice404” email before enter the activation code. After user can download the Zip file and extract that file.

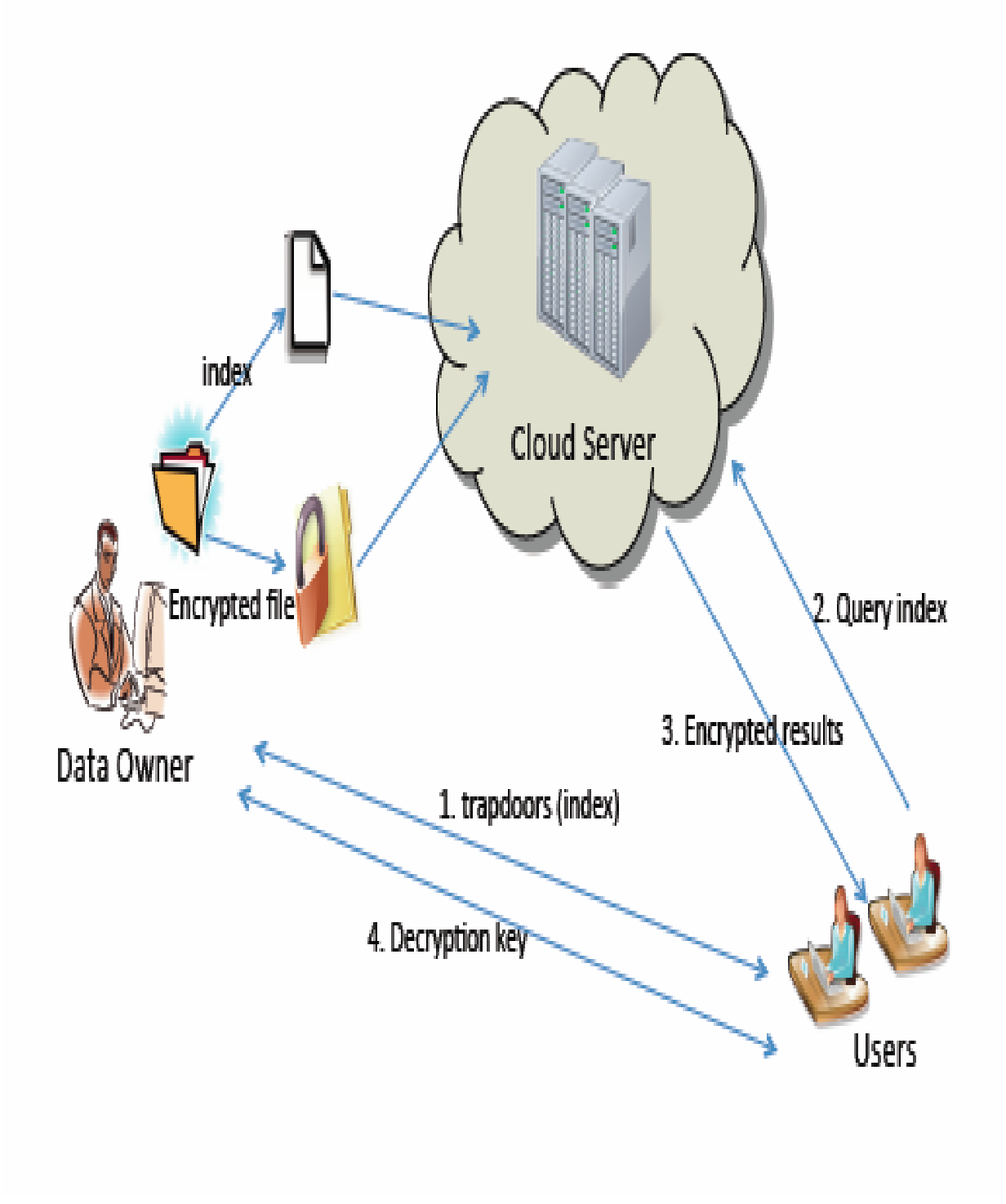
Multi-keyword Module:

This module is used to help the user to get the accurate result based on the multiple keyword concepts. The users can enter the multiple words query, the server is going to split that query into a single word after

search that word file in our database. Finally, display the matched word list from the database and the user gets the file from that list.

Admin Module:

This module is used to help the server to view details and upload files with the security. Admin uses the log key to the login time. Before the admin logout, change the log key. The admin can change the password after the login and view the user downloading details and the counting of file request details on flowchart. The admin can upload the file after the conversion of the Zip file format



FUTURE SCOPE

As our future work, we will explore supporting other multi keyword semantics (e.g.,weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in more stronger threat model. Following the current research, there are possible improvements and undergoing efforts that will appear in the future work. Firstly, the user side of proposed system will be implemented on mobile devices running Android and iOS operating systems since the potential application scenario envisions that users access the data anywhere and anytime. And secondly, the proposed method will be tested on a real dataset in order to compare the performance of our ranking method with the ranking methods used in plain datasets that do not involve any security or privacy-preserving techniques.

CONCLUSION

In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching", i.e., as many matches as possible, to effectively capture similarity between query keywords and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity measurement. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we first propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead on both computation and communication.

REFERENCES

[1] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, 2009.
[2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in RLCPS, January 2010, LNCS. Springer, Heidelberg.
[3] A. Singhal, "Modern information retrieval: A brief overview," IEEE Data Engineering Bulletin, vol. 24, 2001.
[4] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.
[5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches
on encrypted data," in Proc. of S&P, 2000.
[6] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, 2003, http:// eprint.iacr.org/2003/216.
[7] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.
[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.
[9] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key
encryption with keyword search," in Proc. of EUROCRYPT, 2004.
[10] M. Bellare, A. Boldyreva, and A. ONeill, "Deterministic and efficiently searchable encryption," in Proc. of CRYPTO, 2007.