A Novel Approach of Providing Privacy to Encrypted Cloud Data by Using Multi Keywords

GunaSravan Kompalli [#1], **Y.Ramesh Kumar** [*2]
M.Tech Scholar [#1], *HOD & Associate Professor* [*2]
Department of Computer Science & Engineering,
Avanthi Institute of Engineering and Technology, Cherukupally,
Vizianagaram (Dist), AP, India.

Abstract: **In current day's cloud computing has become one of the fascinating domains which are used by almost all MNC and IT companies. Generally this is formed by interconnecting a large number of systems connected all together for remote servers hosted on internet to store, access, retrive data from remote machines not from local machines. As the cloud server has the capability to store a lot of valuable data on its memory block, a lot of users can connect with the centralized location to access, retrieve and modify the data which is stored on the cloud server. Till now there was no mechanism available to store the data in a encrypted manner in all public clouds and even private clouds. So in this paper we have implemented a new concept called as encrypting the data at client side before storing that in the cloud locations. Also as a highest form of security we have integrated a new security primitive like Novel Multi Keywords Ranked Search (NMKRS) over encrypted cloud data in order to give high level of security for the cloud data. By using this NMKRS the data can be downloaded if the user substitutes the series of multi keywords that are placed while data storing by the cloud owner.Initailly the data owner will upload a file which he want to store securely in the cloud with a set of multiple keywords as input, now if the data user want to retrieve or access the same data he need to verify his identify and then download the data with the help of set of assigned keywords. If the user who want to retrieve the data, substitutes wrong identities then the data will not be decrypted and it remains in a encrypted manner. By conducting various experiments on our proposed model, our simulation results clearly tells that this is the first time to implement this type of model to give more level of security for accessing the data to and from remote cloud server.**

Keywords: **Multi Keyword Ranked Search, Encryption, Authentication, Identity Verification.**

1. Introduction

Cloud Computing is one among the fascinating domain in recent analysis space wherever all the info is always processed remotely in unknown machines those users don't own or operate. As cloud computing has raised user attention in storing their valuable knowledge however limits in allocating resources dynamically. The Cloud computing presents a replacement thanks to supplement the present consumption and delivery model for IT services supported the web, by providing for dynamically ascendable and sometimes virtualized resources as a service over the net. Now a days there was a variety of notable business and individual cloud computing services, together with Yahoo, Silicon House, Amazon, Google, Microsoft, and sales division. Moreover, users might not recognize the machines that truly method and host their knowledge [1]. Whereas enjoying the convenience brought by this new technology, users additionally begin worrying regarding

losing management of their own knowledge [2].

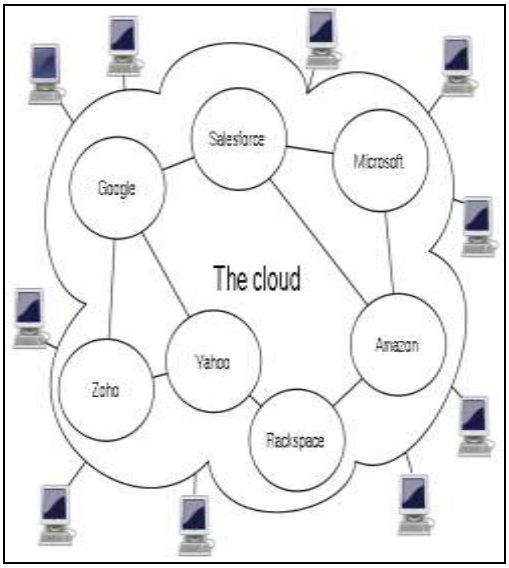

Figure. 1. Represents the Various Cloud Service Providers

From the figure 1,we can clearly represent various types of cloud service providers that are available in current days. All these service providers provide a rich facility to store the valuable data on their individual memory blocks that are provided for the individual cloud users. As we know that a lot of users shown their attention towards the usage of cloud servers for storing their valuable sensitive or private data, there was some limitations that were still not however resolved in cloud server. One of the major limitation that is available in current cloud is data which is stored in the cloud servers either public or private memory blocks is stored in the form of plain text not in any encrypted manner. So this leads a major problem in achieving data integrity for the cloud users. To resolve this problem we have implemented a facility like encrypting the data from the client side before storing the data into the cloud server. By doing this we can able to resolve the first limitation that is available in current clouds.

Another major limitation that was not yet implemented in the current clouds is there was no facility like multi keywords based authentication of data during search process. As we know that data which is stored in the cloud server can be accessed by all the users who are available within their server area, there is chance of misuse of data by un-authorized users who resides within our group of data users. This may lead to modification or updating of the original data without knowing to the data owner. So this is a major limitation that is been faced by current cloud users. So in this paper we have implemented for the first time a new concept called as Multi Keyword Ranked Search over encrypted cloud data, so that the data owner will specify a set of more than one keyword as a security while uploading the data into the cloud server and which is in turn used for search the data by users. The user who substitutes all the series of keys correctly can only be downloading the data and view the data in plain text manner. If the user who try to give wrong keyword during data accessing the data will be remained in encrypted manner and it can't be downloaded for that appropriate user.

2. Related Work

In this section we will find the related work that was analyzed and studied in order to implement this current paper. This section will describe the work related to cloud data storage and search and also the preliminaries that are used in current paper.

A) Preliminary Knowledge

Consider a cloud data hosting service that involves three main entities:

1. Data Owner,
2. Cloud Server and
3. Data User.

The data owner may be an individual or an enterprise, who wishes to outsource a collection of documents D = (D1, D2, . . . , Dn) in encrypted form C = (C1, C2, . . . , Cn) to the cloud server and still preserve the search functionality on outsourced data.

Ci = ES[Di] is the encrypted version of the document Di computed using a semantically secure encryption scheme E with a secret key S. To enable multi-keyword ranked search capability, the data owner constructs searchable index I that is built on m distinct keywords K = (k1, k2, . . . , km) extracted from the original dataset D. Both I and C are outsourced to the cloud server. To securely search the document collection for one or more keywords $K^{-} \in K$, the authorized data user uses search trapdoor (distributed by the data owner) that generates the search request to the cloud server. Once the cloud server receives such request, it performs a search based on the stored index I and returns a ranked list of encrypted documents $L \subseteq C$ to the data user. The data user then uses the secret key S, securely obtained from the data owner, to decrypt received documents L to original view.

We assume an honest-but-curious model for the cloud server. The cloud server is honest, that is, it is always available to the data user and it correctly follows the designated protocol specification, and it provides all services that are expected. The curious cloud server may try to perform some additional analysis to breach the confidentiality of the stored data. In the rest of the paper, the cloud server and the adversary are the same entity. That way, the adversary has access to the same set of information as the cloud server. For this work, we are not concerned about the cloud server being able to link a query to a specific user; nor are we concerned about any denial-of-service attacks.

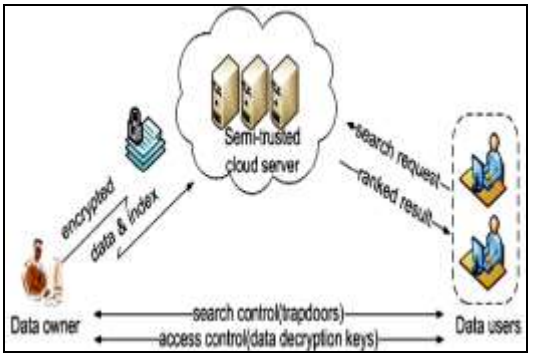

Figure. 2. Represents the Architecture of Search over Encrypted Cloud Data

B) Notations of Proposed Model

Let D = (D1, D2, . . . , Dn) be a set of documents and K = (k1, k2, . . . , km) be the dictionary consisting of unique keywords in all documents in D, where $\forall$ i $\in$ [1,m] ki $\in$ {0, 1} $*$. C = {C1, C2, . . . , Cn} is an encrypted document collection stored in the cloud server. Ii is a searchable index associated with the corresponding encrypted document Ci. If A is an algorithm then a $\leftarrow$ A (. . .) represents the result of applying the algorithm A to given arguments. Let R be an operational ring, we write vectors in bold, e.g. v $\in$ R. The notation v[i] refers to the i-th coefficient of v. We denote the dot product of u, v $\in$ R as u $\otimes$ v = P i=1 u[i] $\cdot$ v[i] $\in$ R. We use |x| to indicate rounding x to the nearest integer, and |x|, |x| (for x > 0) to indicate rounding down or up.

3. Security Using Cryptography Techniques

Cryptography is that the study of techniques for secure communication inside the presence of third parties (called hackers/intruders) lots of sometimes, it's relating to constructing and

analyzing protocols that block adversaries[5], [6] varied aspects in knowledge security like data confidentiality, data integrity, authentication, and non-repudiation are central to modern cryptography. Current cryptography exists at the intersection of the disciplines of arithmetic, technology, and technology. Applications of cryptography embrace ATM cards, laptop passwords, and electronic commerce. Cryptography before the fashionable age was effectively similar with cryptography, the conversion of information from a decipherable state to apparent nonsense. The mastermind of associate degree encrypted message shared the secret writing technique needed to recover the primary knowledge exclusively with meant recipients, thereby precluding unwanted persons from doing the same. Since warfare I and thus the appearance of the laptop, the methods accustomed perform cryptography became increasingly advanced and its application a lot of widespread.

The Modern cryptography is heavily supported mathematical theory and technology practice; crypto logical algorithms are designed around machine hardness assumptions, making such algorithms arduous to interrupt in apply by any somebody. It's in theory potential to interrupt such a system, but it's impossible to undertake and do therefore by any superb smart means. These schemes are therefore termed computationally secure; theoretical advances, e.g., enhancements in range resolution algorithms, and faster computing technology would like these solutions to be often tailored. There exist information-theoretically secure schemes that incontrovertibly cannot be broken even with unlimited computing power—an example is that the one-time pad—but these schemes are harder to implement than the foremost effective in theory breakable but computationally secure mechanisms.

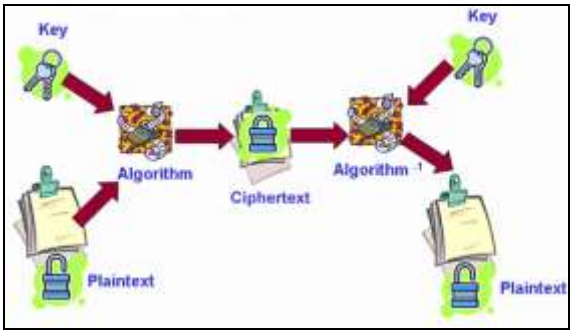

Figure. 3. Represents the Process of Encryption and Decryption of Valuable or Sensitive Data

From the above figure 3, we can clearly find out that for encrypting a plain text we need a key based on the type of algorithm and once the data is encrypted with the help of a key this will be send to the receiver .The receiver will receive the encrypted data and he now try to decrypt the data with the key that was specified by the sender and then finally view the data in the form of plain text.

4. Proposed Novel Multi Keywords Ranked Search (NMKRS): For easy presentation, operations on the data documents are not shown in the framework since the data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data. With focus on the index and query, the NMKRS system consists of four algorithms as follows:

- Setup (1^{ℓ}). *Taking a security parameter ℓ as input, the data owner outputs a symmetric key as SK.*
- BuildIndex ($\mathcal{F}, SK$). *Based on the data set $\mathcal{F}$, the data owner builds a searchable index $\mathcal{I}$ which is encrypted by the symmetric key SK and then outsourced to the cloud server. After the index construction, the document collection can be independently encrypted and outsourced.*
- Trapdoor ($\widetilde{W}$). *With t keywords of interest in $\widetilde{W}$ as input, this algorithm generates a corresponding trapdoor $T_{\widetilde{W}}$.*
- Query ($T_{\widetilde{W}}, k, \mathcal{I}$). *When the cloud server receives a query request as ($T_{\widetilde{W}}$, k), it performs the ranked search on the index $\mathcal{I}$ with the help of trapdoor $T_{\widetilde{W}}$, and finally returns $\mathcal{F}_{\widetilde{W}}$, the ranked id list of top-k documents sorted by their similarity with $\widetilde{W}$.*

Neither the search control nor the access control is within the scope of this paper. While the former is to regulate how authorized users acquire trapdoors, the later is to manage users' access to outsourced documents.

Keyword Privacy

As users usually prefer to keep their search from being exposed to others like the cloud server, the most important concern is to hide what they are searching, i.e., the keywords indicated by the corresponding trapdoor. Although the trapdoor can be generated in a cryptographic way to protect the query keywords, the cloud server could do some statistical analysis over the search result to make an estimate. As a kind of statistical information, document frequency (i.e., the number of documents containing the keyword) is sufficient to identify the keyword with high probability [31]. When the cloud server knows some background information of the data set, this keyword specific information may be utilized to reverse engineer the keyword.

5. Implementation Modules

Implementation is the stage where theoretical design is converted into programmatically way. Generally in the implementation stage we will divide the application into number of modules in order to make the application develop very easily. We have implemented the proposed concept on Java Platform in order to show the performance this proposed multi keyword ranked based search over encrypted cloud data. The front end of the application takes JSP, HTML and Back-end takes My SQL Server 5.0 along with a Real Cloud Service provider called as DRIVEHQ Cloud Service provider. This cloud service provider will provide a space up to 2 GB for storing the files which is used by the application. The application is divided mainly into following 4 modules. They are as follows:

A. Data Owner Module
B. Data User Module
C. Encryption Module
D. Rank Search Module

A) Data Owner Module

This module helps the owner to register those details and also include login details. This module helps the owner to upload his file with encryption using RSA algorithm. This ensures the files to be protected from unauthorized user.

B) Data User Module

This module includes the user registration login details. 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 email before enter the activation code. After user can download the Zip file and extract that file.

C) Encryption Module

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

D) Rank Search Module

These modules ensure the user to search the files that are searched frequently using rank search. This module allows the user to download the file using his secret key to decrypt the downloaded data. This module allows the Owner to view the uploaded files and downloaded files

.

6. 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 similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. Also as an extension we have implemented the proposed model on real cloud like "DRIVEHQ" in order to store the data and retrieve the data to and from the server. By implementing this real cloud the data which is uploaded through our application will be directly saved into the drivehq cloud server in the specified user account.

References

[1] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. Sedic: privacyaware data intensive computing on hybrid clouds. In *Proceedings of the 18th ACM conference on Computer and communications security*, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM

[2] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*,22(1):1–61, 2009.

[3] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.

[4] E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, http://eprint.iacr.org/2003/216. 2003.

[5] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,"Proc. 13th ACM Conf. Computer and Comm. Security (CCS '06), 2006.

[6] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data," Proc. Third Int'l Conf. Applied Cryptography and Network Security, 2005.

[7] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), 2004.

[8] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Cryptography from Anonymity," Proc. IEEE 47th Ann. Symp. Foundations of CS, pp. 239-248, 2006.

[9] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-June 2012.

[10] J. Zobel and A. Moffat, "Exploring the Similarity Space," ACM SIGIR Forum, vol. 32, pp. 18-34, 1998.

About the Authors

GunaSravan Kompalli is currently pursuing his 2 years M.Tech in CSE Avanthi Institute of Engineering and Technology, Cherukupally, Vizianagaram (Dist), AP, India His area of interest includes Networks and Cloud Computing.

Y.Ramesh Kumar has completed M.Tech (SE) in Andhra University. He has more than 10 years of experience in teaching field. Presently he is working as an Associate Professor & HOD in Department of Computer Science & Engineering Avanthi Institute of Engineering and Technology, Cherukupally, Vizianagaram District, Andhra Pradesh.His resarch interest includes Web Technologies, OOPS,Networks,Mining,Security.

