PRMSM : A NOVEL PROTOCOL FOR ANALYZING THE PERFORMANCE OF A RANKED SEARCH OVER A SECURE MULTI - CLOUD DATA STORAGE

BALLA NAGA MANOJ [#1] , A.V.S.PAVAN KUMAR [#2]

[#1] M.Tech Scholar,Department of Computer Science and Technology,
Baba Institute of Technology and Sciences,Visakhapatnam,AP,India.

[#2] Assistant Professor,Department of Computer Science and Engineering,
Baba Institute of Technology and Sciences,Visakhapatnam,AP,India.

ABSTRACT

In present days cloud has achieved a great importance mainly in terms of storage and retrieval of data from remote systems rather than from its local machine. As data is stored on the remote systems, it will be accessed remotely via internet by connecting local system with its main server. The main limitation of the current cloud servers is data is stored in plain text rather than in an encrypted manner. As the data is stored in this form there is no security for the sensitive data which is stored in the cloud servers. As we all know that some clouds may use any of the encryption algorithms for encrypting the data before it is stored into the cloud server, it has some limitations when compared one with another in its individual functionality. There are mainly two limitations in the current cloud service providers where the first limitation is all the data which is stored on the cloud server is stored in the normal manner or in plain text so that it can be viewed and modified by anyone within the group. Also we know that in current cloud servers there is no concept like multiple keywords for search, there is only single keyword search, which leads a great difficulty for users to search the given data with exact keyword what they gave during uploading time. In this thesis, we have introduced a novel concept like data encryption before it is stored directly into the cloud server like DRIVEHQ service what we used in our current application. In this paper as an extension we have also implemented a new concept like advanced authorization of cloud users, where the cloud server need to give activation permission for the registered users or owners. Those who get the permission after registration will receive the login password for their registered mail id,with that only the user or owner can login, if not login fails this give more security for the current application. By conducting various experiments on our proposed application by using the DRIVEHQ as the back end storage service for the current application, we finally came to a conclusion that this is the first time to implement such a function which was not yet implemented in any of the primitive cloud service providers for data during insertion and retrieval compared to various primitive clouds.

Key Words: Ranked Search, Additive Order, Encryption, Multi Keyword Search.

I. INTRODUCTION

In current days cloud computing domain has occupied a major role in each and every part of the information processing and information storage centers. As the cloud has became a valuable resource for all parts of information processing centers, the data which is to be stored will be stored on the remote systems not on their local hardware, and accessed remotely via internet by connecting various servers. As the data will be stored on remote server, the data user need to retrieve the data from the remote server, whenever he want any data from that remote hardware. In the current cloud servers, the major limitation is data which is stored and shared over the cloud users has no security and there is also no security for accessing the data in the current cloud servers [1]. This is mainly because all the data which is stored in the current cloud servers is stored in the form of plain text rather than in a cipher text manner. As we know that cloud has exaggerated user attention in storing their valuable or sensitive information however limits in allocating resources dynamically. As we know that cloud has received more and more user's attention towards data storage, it still has some restrictions in size constraints. In enterprise settings, we tend to see the increase in demand for knowledge outsourcing that assists within the strategic management of corporate knowledge. In the recent cloud service providers, it is straightforward to use without charge accounts for email, image album, file sharing and/or remote access, with storage size a lot of than Fifteen GB (for free usage) and up to 1 TB or more for the premium users [2].Next in the current cloud service providers there is no concept like ranking the files which is stored and uploaded by the data owners. In the cloud there are various types of services available in which Data Base as a Service (DaaS) is one of the main and prominent services among others. This service is not having security for the data which is stored in the cloud, compared with various other cloud services, hence our main motto is to provide security for this DaaS service by integrating various encryption and other techniques are proposed in this current paper.

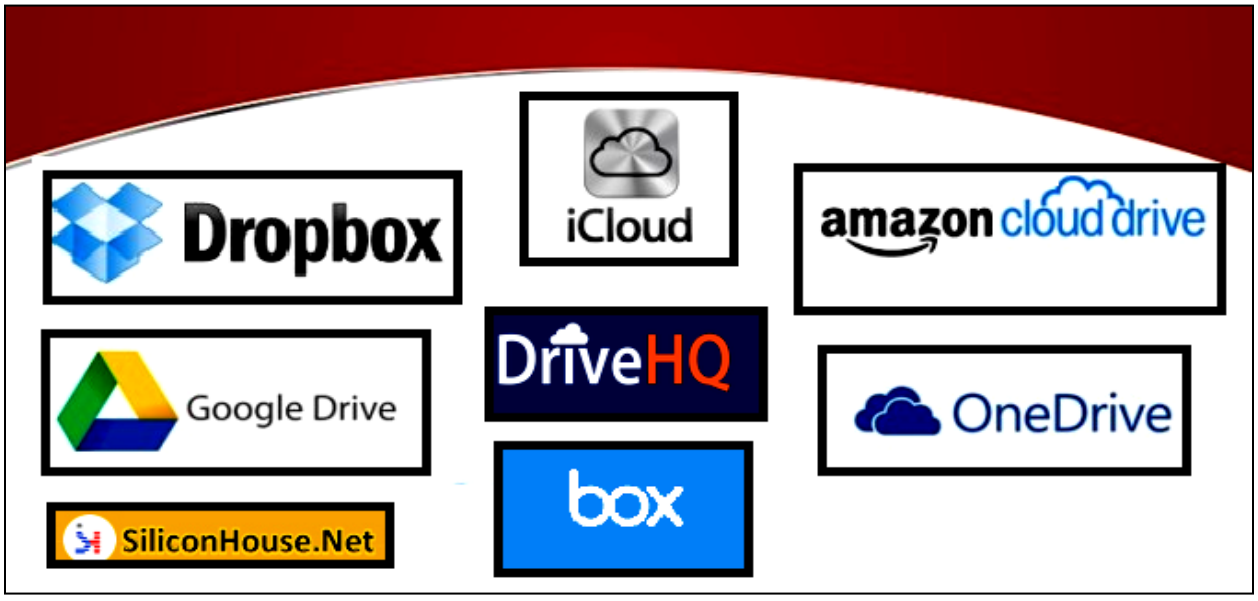

Figure.1. Various types of Real Time Cloud Service Providers

From the above figure 1, we can clearly find out that there are many cloud service providers that are available in the current days for storing and accessing the data remotely. Here in the above figure, there are some public clouds which takes no amount for storing the data till 2GB,some clouds are their which will give access only for the premium members like those who have premium account. Also there are some cloud service providers like hybrid cloud, which can provide public and private services at a time also known as hybrid cloud service provider. For our project we are using the **DRIVEHQ** as the cloud service provider for storing and accessing the data in a secure manner .In this project we take a DRIVEHQ account like "**BABACLOUD16**" which is a public cloud which can accept any data up to maximum of 1 GB, which is almost very high and if the data exceeds then it will ask to pay the amt greater than 15 GB for the cloud data user [3], [4].

II. RELATED WORK

In this section we mainly discuss about the various cloud services that are available and also the detailed explanation about each and every service. Generally in the cloud there are mainly four types of services like: IaaS, PaaS, SaaS, and Daas

From the below figure 2, we can clearly find out that there are four different services available and one among them is DaaS, which is the main service what we are using now for providing security for that and prove that this service also gives the best security for the data which is stored inside the cloud memory locations [5], [6]. Now let us discuss about each and every service in detail as follows:

A. IaaS (Infrastructure as a Service)
B. PaaS(Platform as a Service)
C. SaaS(Software as a Service)
D. DaaS (Data /Data Base as a Service)

A. IaaS (Infrastructure as a Service)

In this service the cloud server mainly deals with application level and it is basically used to set the platform for the users. The main persons who come under this service is IT Professionals, this is clearly shown in the figure 2.

B. PaaS (Platform as a Service)

The second and one of the most important service in cloud computing is Platform as a Service, where this is mainly used for customization of cloud server, where the developer comes under this service. Here the cloud server customizes which type of platforms is needed for their company usage is seen in this service.

C. SaaS (Software as a Service)

The third and one of the best services in cloud computing is Software as a Service, where this is mainly used for a consumer to use the cloud service provider's applications running on a

cloud IaaS. Generally business end-users come under this service where all the software's that are required for running the cloud are processed in this service.

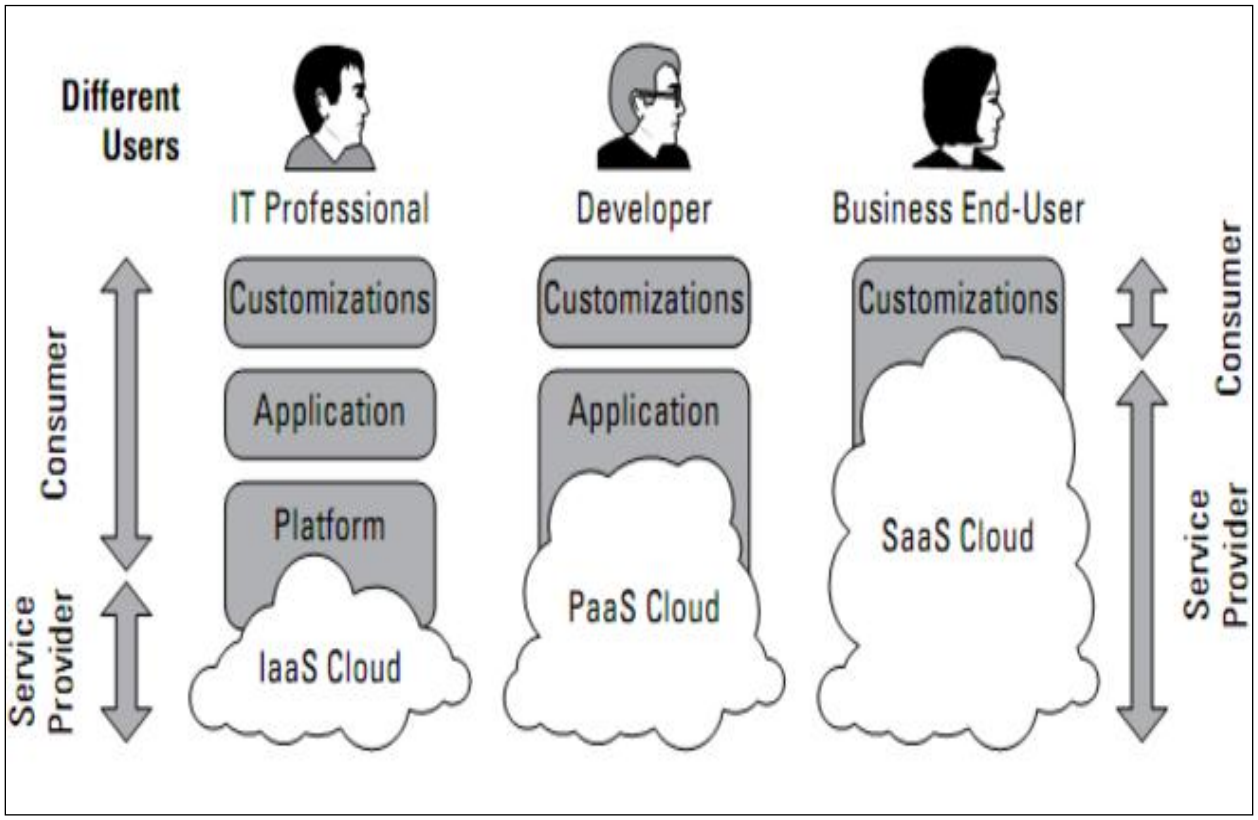

Figure.2. Functionality of Various Cloud Services in Real Time

D. DaaS (Data/Database as a Service)

The last and one of the new services that was launched and included in various cloud client services is DaaS, which is clearly seen in figure 2.This DaaS service is used mainly for storing the data base, tables and data in the form of fragments and packets [7],[8],[9]. As this is having various advantages compared with other cloud client services, it has a small limitation like the data which is stored in this DaaS is not stored in the encrypted manner which is stored in the plain manner.

III. A NOVEL PROPOSED PRIVACY PRESERVING RANKED MULTI-KEYWORD SEARCH IN A MULTI-OWNER MODEL (PRMSM)

In this section we will find out the Novel proposed privacy preserving ranked multi keyword search in a multi owner model and its architecture that was used in the current paper. In this paper we have implemented privacy preserving multi keyword search in a multi owner model also known as PRMSM.Now let us discuss about this in detail as follows:

PRELIMINARIES

Initially we will try to find out the preliminaries that were used in the current paper for performing the application with proposed algorithm. Generally there are four important entities to verify the proposed algorithm, now let us discuss about that in detail

1. Cloud Server,
2. Data Owner
3. Data User
4. Admin

The Cloud Server is the primary entity in the current application where the cloud server will store all the sensitive information that was uploaded by the data owner for giving access for the end users. The data owner entity is the starting entity which is defined as a person who may be an individual or sometimes 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. Here we assume that documents are labeled with D and if there are many documents to be out sourced they are represented as D1, D2 and so on.Here in our proposed application, we take sample text documents as input where initially all the text documents are of plain text and our main motto is to store them in a secure manner inside a cloud server. Once the text files which consists the sensitive data are encrypted and then they are stored into the cloud server,they are termed as C1, C2 and so as they were encrypted by the data owner at his level before out sourcing into the Admin.

Initially the admin enters into the account and once he enters, he has the facility to receive the data request which is send to that by data owner after an initial encryption. This admin will now receive all the files which were uploaded by various data owners and then they will be re-encrypted by the admin at his level and then it was send to the cloud server. During this stage the re-encrypt method will encrypt not only the file content but also the file details like file name,file upload date and time and so on.This double encryption or re-encrypt gives much more security for our proposed application compared with various primitive cloud service providers. Now the data which is uploaded by admin will be reached to the cloud server, where the cloud server is an important entity among all the four as this is the only entity which has the capability to store the encrypted documents into its storage area. Once the cloud owner encrypts the text documents and it is uploaded [9]- [12], then immediately they will be received by the cloud server and it will then store in its storage area securely. When a search user try to download any file, he will send the input as either filename or file keyword so that immediately the file request will be identified by the data base records and if the input keyword is matched the file will be downloaded and if that was not matched it will be identified as data not found. During this process if any intruder try to access illegally the data by substituting the others identies, he will be identified as a trapdoor user and file can’t be downloaded.

Cloud Data user is the last entity who wishes to download the files from the cloud server by giving valid inputs for searching the files and then download those files in a secure manner. The search user has following three steps to be performed for downloading the Encrypted text documents from the cloud server, they are as follows: First, the search user initially after

registration, he will be login into his account by substituting all the valid details what he stored during registration. Second, according to the search keywords, the search user uses the same secret key along with any of the search parameter to generate a decryption key and sends it to the cloud server. Once if the input parameters along with secret key are matched with server records, then the search user receives the matching document collection from the cloud server and decrypts them with the symmetric key which is dynamically sent to the search user mail id at the time of downloading the file. If the user substitutes valid decryption key can only download the file in a plain manner or else the file will be in encrypted manner.

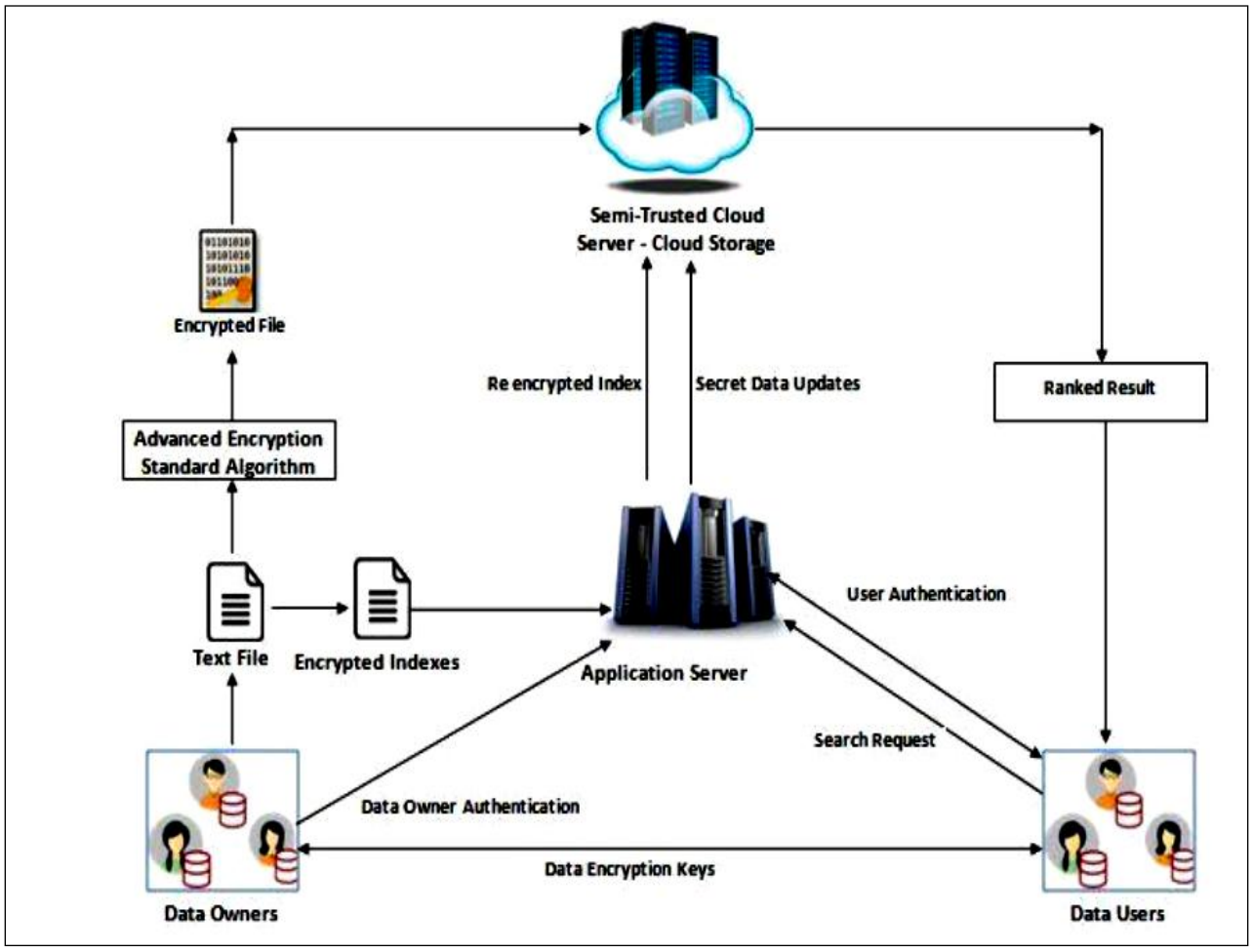

Figure 3. Represents the Architecture of Our Proposed Model

The above diagram clearly represent that there are mainly four roles like Data owner, Cloud Server, Admin Server and Search Users. Here each and every one has individual roles and all the access in this current model is in form of De-Centralized manner. In the primitive or existing cloud servers, the data access will be obviously in a centralized manner, where the data which is uploaded by owner will be stored inside the cloud server and in turn the access will be in the hands of server itself, But there was no single access control for the owner or user in the current cloud service providers. So in this paper we for the first time implemented architecture like De-Centralized access by giving individual rights for each and every individual. The data

which is uploaded by the data owner is having a right to give access or deny the access of his uploaded file at the time of user search request. Here the owner will receive all the search requests done by various data users or search users within the cloud and once if the data owner really wish to give access to user then only he will click on allow button so that access will be granted and symmetric key as a decryption key will be send for the requested search user, if not access will be restricted by the owner and he will be treated him as a trapdoor user. Here the Cloud server has a capability to receive all the user requests and in turn send that request to the appropriate data owners who uploaded the data into the cloud. Here the data which is uploaded is in form of encrypted manner and the records are almost text documents with a valid sensitive data and they are stored in a secure manner onto the cloud storage area[14].

Assumptions

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.

Here in the current application we denote the function Ci = E_S[Di] is the encrypted version of the document D_i, which is mainly computed by using a semantically secure encryption scheme E with a secret key S. To enable multi-keyword ranked search capability, the data owner always constructs a searchable index termed as "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.

IV. IMPLEMENTATION PHASE

Implementation is the stage where the theoretical design is converted into programmatically manner.In this stage we will divide the application into a number of modules and then coded for deployment. We have implemented the proposed concept on Java programming language with JEE as the chosen language in order to show the performance this proposed Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM). The front end of the application takes JSP, HTML and Java Beans and as a Back-End Data base we took My SQL data base along with a Real Cloud Service provider called as DRIVEHQ Cloud Service provider .For this we have created a public cloud account in DRIVEHQ as "**BABACLOUD16"** with a maximum space up to 1 GB for storing the files which is used by the application. The proposed application is divided mainly into following four modules. They are as follows:

1. Network Construction
2. Cloud User Authentication
3. Multi Owner Search
4. Illegal Search Identification

1. NETWORK CONSTRUCTION

This is the first module in which we will try to create a network to implement our current proposed system. This module contains mainly the following roles like admin, data users, data owners, and a single cloud server. Admin provides the accessibility to Data-owners. Initially Data-owner needs to register and admin approves the each data owner request. The respective Password and login credentials will be sent to the Email ID of Data owner. In Users sub-module, each user has a global identity in the system. A user may be entitled a set of attributes which may come from multiple attribute authorities. The user will receive a secret key associated with its attributes entitled by the corresponding attribute authorities. In data owner's sub-module, the proposed scheme should allow new data owners to enter this system without affecting other data owners or data users, i.e., the scheme should support data owner scalability in a plug-and-play model. In Cloud Server sub-module of system model, the owner sends the encrypted data to the cloud server through Admin. They do not rely on the server to do data access control. But, the access control happens inside the cryptography. That is only when the user's attributes satisfy the access policy defined in the cipher text; the user is able to decrypt the cipher text. Thus, users with different attributes can decrypt different number of content keys and thus obtain different granularities of information from the same data.

2. CLOUD USER AUTHENTICATION

Here as in order to identify whether the user who is accessing the data is valid or invalid, each and every data user should be authenticated if he /she is valid or invalid. To prevent attackers from pretending to be legal data users performing searches and launching statistical

attacks based on the search result, data users must be authenticated before the administration server re-encrypts trapdoors for data users. Traditional authentication methods often follow three steps. First, data requester and data authenticator share a secret key, say, *k*0. Second, the requester encrypts his personally identifiable information *d*0 using *k*0 and sends the encrypted data (*d*0) *k0* to the authenticator. Third, the authenticator decrypts the received data with *k*0 and authenticates the decrypted data. The key point of a successful authentication is to provide both the dynamically changing secret keys and the historical data of the corresponding data user.

3. MULTI OWNER SEARCH DETECTION

In this module we have a facility to encrypt the data with different keys for different data owners. It also needs to allow the cloud server to rank the search results among different data owners and return the top-*k* results. The cloud server stores all encrypted files and keywords of different data owners. The administration server will also store a secret data on the cloud server. Upon receiving a query request, the cloud will search over the data of all these data owners. The cloud processes the search request in two steps. First, the cloud matches the queried keywords from all keywords stored on it, and it gets a candidate file set. Second, the cloud ranks files in the candidate file set and finds the most top-*k* relevant files. Finally, we apply the proposed scheme to encode the relevance scores and obtain the top-*k* search results.

4. ILLEGAL SEARCH IDENTIFICATION

This is one of the main and crucial module among all modules in our application where the authentication process is protected by the dynamic secret key and the historical information. We assume that an attacker has successfully eavesdropped the secret key. Then he has to construct the authentication data; if the attacker has not successfully eavesdropped the historical data, e.g., the request counter, the last request time, he cannot construct the correct authentication data. Therefore this illegal action will soon be detected by the administration server. Further, if the attacker has successfully eavesdropped all data of *Uj*, the attacker can correctly construct the authentication data and pretend himself to be *Uj* without being detected by the administration server.

V. CONCLUSION

In this paper, we for the first time have implemented a secure multi-keyword search for multiple data owners and multiple data users in the cloud computing environment. In the current cloud servers, there was no concept like encryption of data before it is stored into the cloud. As the encryption was not available in the current clouds, all the data which is stored into the cloud has no security and any one can access that data freely without any restrictions. So in this paper for the first time we have implemented a new concept like encryption of data before it is stored inside the live cloud. Also we have implemented a new concept called as multi keyword search, where the data which is uploaded by the data owner will provide multiple attributes for each and every file during the upload, so this multiple attributes act like a multiple keywords for accessing the file during download or search. By conducting various experiments on our proposed model, we finally came to a conclusion that this proposed mechanism gives high level of security in terms of data during storage and retrieval.

VI. REFERENCES

[1] C. Wang and W. Lou, “A New Privacy- protective Public Auditing for Secure Cloud Storage,” IEEE Trans.Computers, vol. 62, Feb. 2013.

[2] G.C. Chick and S.E. Tavares, “Flexible Access management with Master Keys,” Proc. Advances in cryptography (CRYPTO ’89), vol. 435, pp. 316-322, 1989.

[3] "About Dropbox". *Dropbox, Inc*. Retrieved 2013-06-03. *Dropbox was founded by Drew Houston and Arash Ferdowsi in 2007, and received seed funding from Y Combinator*.

[4] "Meet the Team! (Part 1)". *The Dropbox Blog. Dropbox, Inc*. Retrieved April 24, 2010 by *Ying, Jon (February 5, 2009).*.

[5] Peter Mell and Timothy Grance (September 2011). The NIST Definition of Cloud Computing (Technical report). National Institute of Standards and Technology: U.S. Department of Commerce. doi:10.6028/NIST.SP.800-145. Special publication 800-145.

[6] Alcaraz Calero, Jose M.; König, Benjamin; Kirschnick, Johannes (2012). "Cross-Layer Monitoring in Cloud Computing". In Rashvand, Habib F.; Kavian, Yousef S. Using Cross-Layer Techniques for Communication Systems. Premier reference source.

[7] IGI Global. p. 329. ISBN 978-1-4666-0961-7. Retrieved 2015-07-29. Cloud Computing provides services on a stack composed of three service layers (Hurwitz, Bloor, Kaufman, & Halper, 2009): Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS).

[8] Gartner; Massimo Pezzini; Paolo Malinverno; Eric Thoo. "Gartner Reference Model for Integration PaaS". Retrieved 16 January 2013.

[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] 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.

[11] http://www.merriam-webster.com/dictionary/ranking.

[12] Towell, G.G. and Shavlik, J.W. (1993), ‘Extracting refined rules from knowledge-based neural networks’ physicist. Learn, Vol.13, pp.71-101.

[13] Ehrlich, Melanie; Gama-Sosa, Miguel A.; Huang, Lan-Hsiang; Midgett, Rose Marie; Kuo, Kenneth C.; McCune, Roy A.; Gehrke, Charles (1982). "Amount and distribution of 5-methylcytosine in human DNA from different types of tissues or cells". Nucleic Acids Research.

[14] Moréra, Solange; Larivière, Laurent; Kurzeck, Jürgen; Aschke-Sonnenborn, Ursula; Freemont, Paul S; Janin, Joël; Rüger, Wolfgang (August 2001).

VII. ABOUT THE AUTHORS

BALLA NAGA MANOJ is currently pursuing his 2 years M.Tech in Department of Computer Science and Technology at Baba Institute of Technology and Sciences, Visakhapatnam, AP, India. His area of interest includes Cloud with Security.

A.V.S. PAVAN KUMAR is currently working as an Assistant Professor in Department of Computer Science and Engineering at Baba Institute of Technology and Sciences, Visakhapatnam, AP, India. He received M.Tech from Gitam University and he has more than 5 years of teaching experience in engineering colleges.His research interest includes Data Mining and Machine Learning.