A PROFICIENT SECURE APPLICABLE-RANKED KEYWORD SEARCH METHOD

PALA SUPHALA [#1] , D.D.D SURIBABU [#2]

[#1] M.Tech Scholar , Department of Computer Science and Engineering,
D.N.R College of Engineering and Technology,
Balusumudi, Bhimavaram, India.

[#2] HOD & Associate Professor , Department of Computer Science and Engineering,
D.N.R College of Engineering and Technology,
Balusumudi, Bhimavaram, 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 in the current cloud server there is no facility like ranked search over encrypted cloud data, which is nothing but showing the top priority accessed files separately than compared with low priority accessed files. Hence in this proposed thesis, we for the first time have implemented a concept like finding the Top K documents that are available from the set of documents that are available inside the cloud server. By conducting various experiments on our proposed hierarchal index model for storing and accessing the data securely to and from the cloud server, our simulation results clearly tells that our proposed method gives best level of security for storing the data and retrieving the data from the cloud server.

Key Words: Ranked Search, Data Encryption, Cloud Service Provider, hierarchal Clustering Model and Encrypted Search.

I. INTRODUCTION

Now a day's cloud computing domain has attained 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.

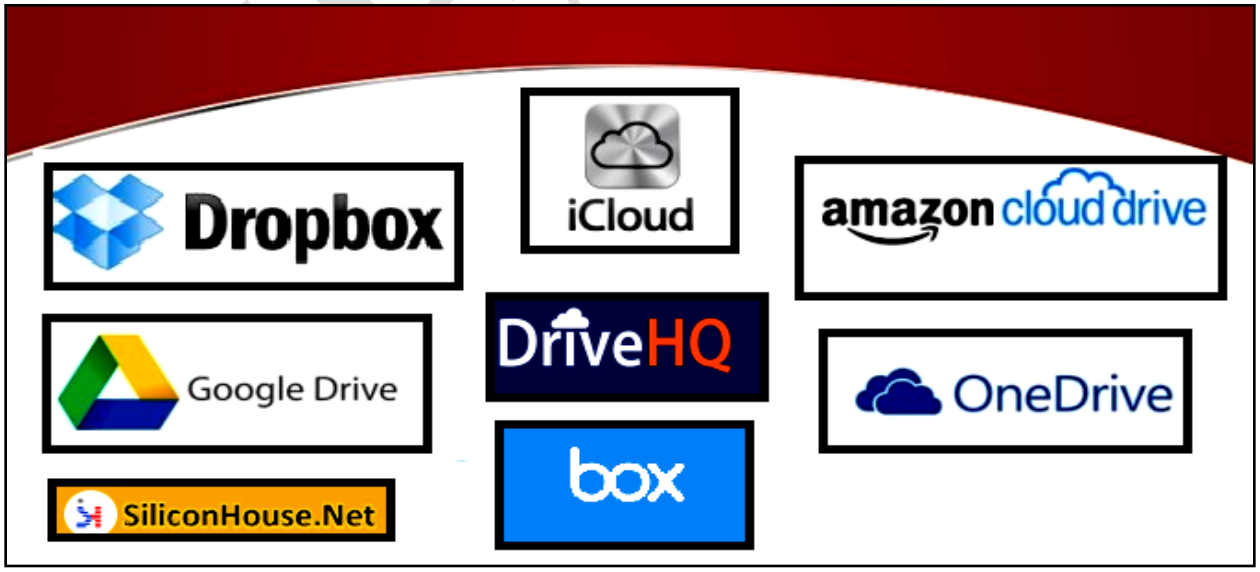

FIGURE.1. REPRESENTS THE DIFFERENT REAL TIME CLOUD SERVICE PROVIDERS

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 [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:

1. IaaS,
2. PaaS,
3. SaaS, and
4. 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)

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.

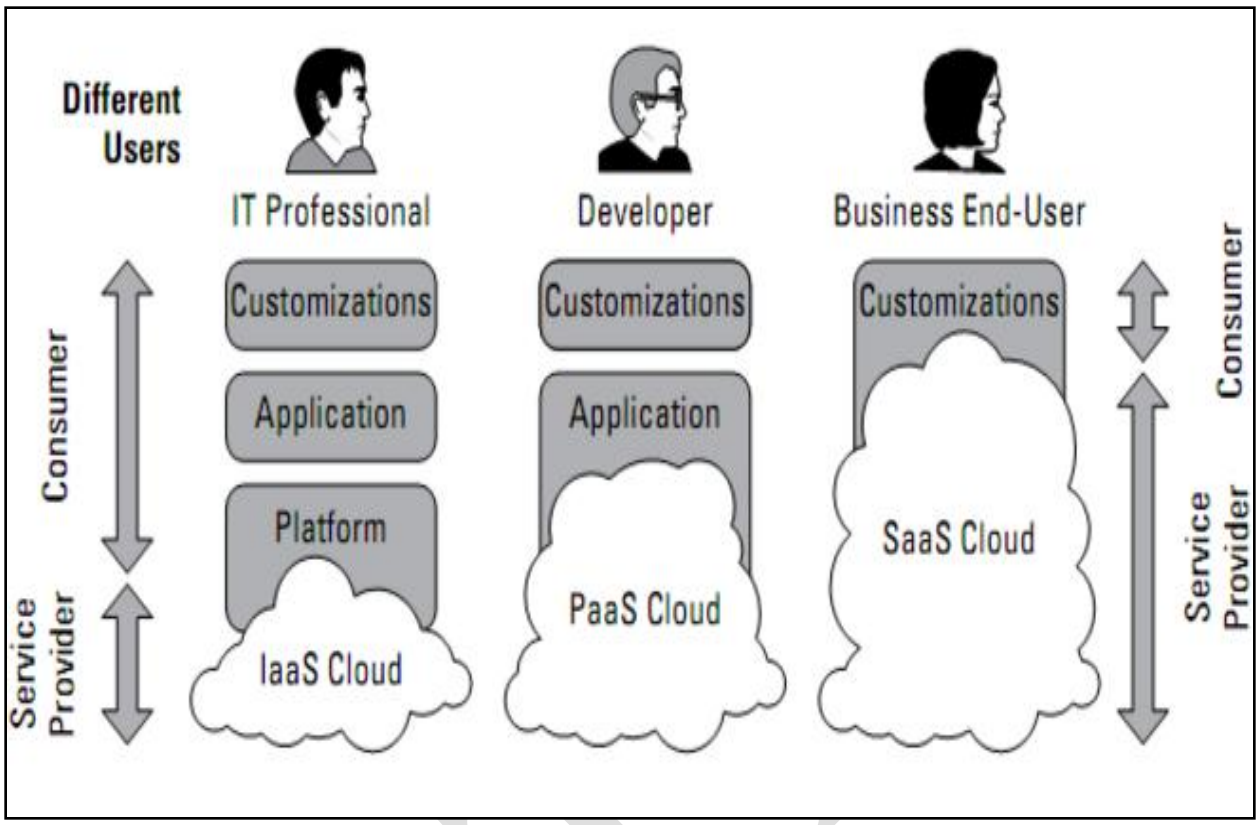

FIGURE.2. REPRESENTS THE INDIVIDUAL FUNCTIONALITY OF VARIOUS REAL TIME CLOUD SERVICES

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. Our current thesis work s mainly concentrated on DaaS by providing security for the data in terms of encryption and also has a control on the search data.

III. A NOVEL PROPOSED MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED DATA BASED ON HIERARCHICAL CLUSTERING INDEX (NMRSE-HCI)

In this section we will find out the proposed novel multi-keyword ranked search over the encrypted data based on hierarchical clustering index protocol and the architecture of the current application is shown clearly in the below architecture. Here we used a hierarchical clustering approach for efficient extraction of top K-documents from the set of documents that are available inside the cloud server. Now let us discuss about this in detail as follows:

PRELIMINARY KNOWLEDGE

In this current thesis, a vector space model is used and every document is represented by a vector, which means every document can be seen as a point in a high dimensional space. Due to the relationship between different documents, all the documents can be divided into several categories. In other words, the points whose distance is short in the high dimensional space can be classified into a specific category. The search time can be largely reduced by selecting the desired category and abandoning the irrelevant categories. Comparing with all the documents in the dataset, the number of documents which user aims at is very small. Due to the small number of the desired documents, a specific category can be further divided into several sub-categories. Instead of using the traditional sequence search method, a backtracking algorithm is produced to search the target documents.

Cloud server will first search the categories and get the minimum desired sub-category. Then the cloud server will select the desired k documents from the minimum desired sub-category. The value of k is previously decided by the user and sent to the cloud server. If current sub-category cannot satisfy the k documents, cloud server will trace back to its parent and select the desired documents from its brother categories. This process will be executed recursively until the desired k documents are satisfied or the root is reached. To verify the integrity of the search result, a verifiable structure based on hash function is constructed. Every document will be hashed and the hash result will be used to represent the document. The hashed results of documents will be hashed again with the category information that these documents belong to and the result will be used to represent the current category. Similarly, every category will be represented by the hash result of the combination of current category information and sub-categories information. A virtual root is constructed to represent all the data and categories. The virtual root is denoted by the hash result of the concatenation of all the categories located in the first level. The virtual root will be signed so that it is verifiable. To verify the search result, user only needs to verify the virtual root, instead of verifying every document.

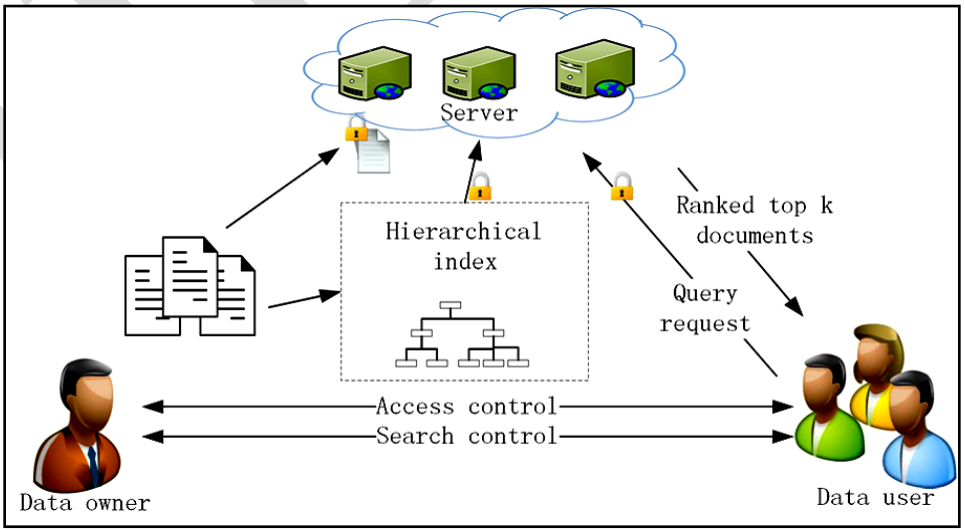

FIGURE.3. REPRESENTS THE ARCHITECTURE OF CIPHERTEXT SEARCH BY THE DATAUSER OVER AN REAL TIME CLOUD SERVER

From the above figure 3, we can clearly find out that in our proposed system model we have three main entities, like the data owner, the data user, and the cloud server. Also we can find out some dashed lines in the figure indicates the added component to the existing architecture.

In any of the cloud server, the data owner is responsible for collecting documents, building document index and outsourcing them in an encrypted format to the cloud server.
Apart from that, the data user needs to get the authorization from the data owner before accessing to the data. The cloud server provides a huge storage space, and the computation resources needed by cipher text search. Upon receiving a legal request from the data user, the cloud server searches the encrypted index, and sends back top-k documents that are most likely to match users query [12]. The number k is properly chosen by the data user. Our system aims at protecting data from leaking information to the cloud server while improving the efficiency of cipher text search. In this model, both the data owner and the data user are trusted, while the cloud server is semi-trusted, which is consistent with the architecture in [10, 11, 12]. In other words, the cloud server will strictly follow the predicated order and try to get more information about the data and the index.

THREAT MODEL

The adversary's ability can be concluded in two threat models.

KNOWN CIPHERTEXT MODEL

In this model, Cloud server can get encrypted document collection, encrypted data index, and encrypted query keywords.

KNOWN BACKGROUND MODEL

In this model, cloud server knows more information than that in known ciphertext model. Statistical background information of dataset, such as the document frequency and term frequency information of a specific keyword, can be used by the cloud server to launch a statistical attack to infer or identify specific keyword in the query [10, 11], which further reveals the plaintext content of documents. The adversary's ability can be represented in the above two threat models.

IV.NMRSE-HCI ARCHITECTURE

In this section we mainly discuss about the NMRSE-HCI architecture that was implemented in our current application. Now let us look about that in detail as follows:

The hierarchical index structure is introduced into the MRSE-HCI instead of sequence index. In MRSE-HCI, every document is indexed by a vector. Every dimension of the vector stands for a keyword and the value represents whether the keyword appears or not in the document. Similarly, the query is also represented by a vector. In the search phase, cloud server

calculates the relevance score between the query and documents by computing the inner product of the query vector and document vectors and return the target documents to user according to the top k relevance score.

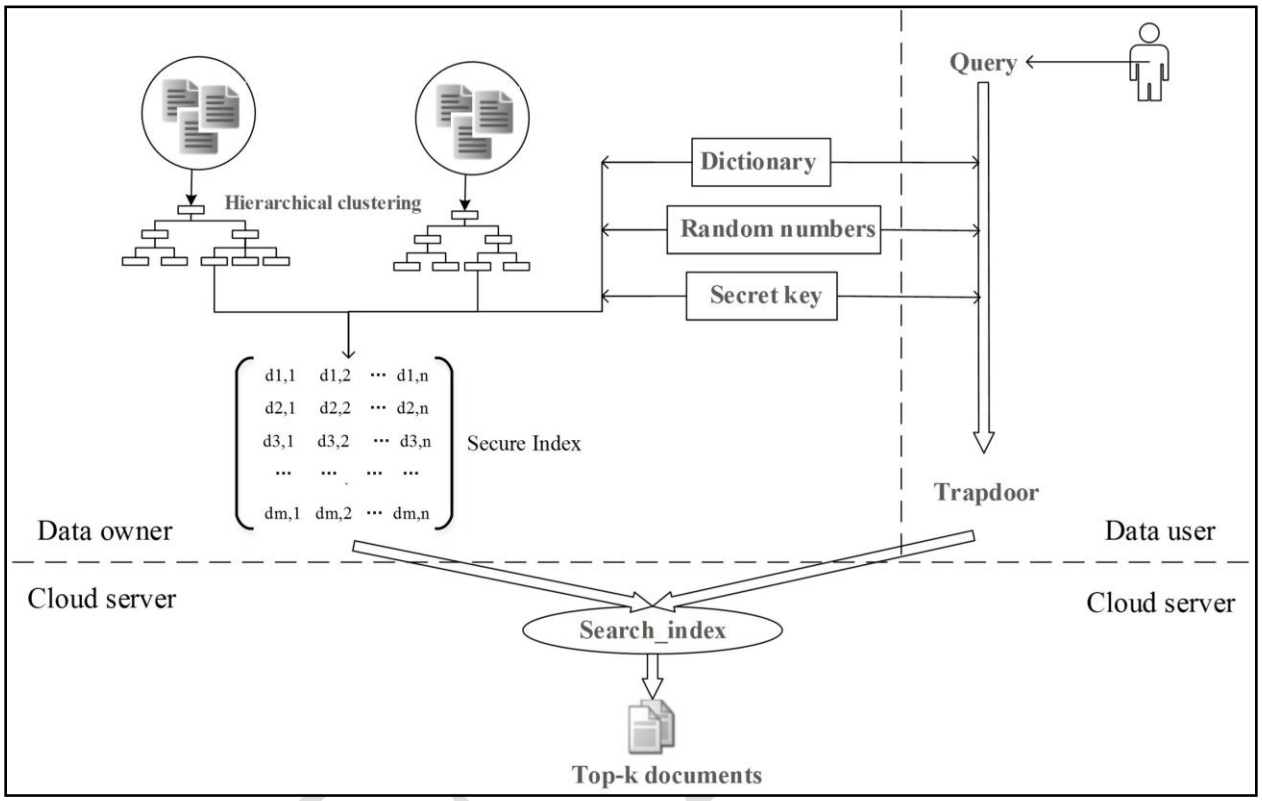

FIGURE.4. REPRESENTS THE ARCHITECTURE OF OUR PROPOSED NMRSE-HCI PROTOCOL

Due to the fact that all the documents outsourced to the cloud server is encrypted, the semantic relationship between plain documents over the encrypted documents is lost. In order to maintain the semantic relationship between plain documents over the encrypted documents, a clustering method is used to cluster the documents by clustering their related index vectors. Every document vector is viewed as a point in the n-dimensional space. With the length of vectors being normalized, we know that the distance of points in the n-dimensional space reflect the relevance of corresponding documents. In other word, points of high relevant documents are very close to each other in the n-dimensional space. As a result, we can cluster the documents based on the distance measure. With the volume of data in the data center has experienced a dramatic growth, conventional sequence search approach will be very inefficient. To promote the search efficiency, a hierarchical clustering method is proposed. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold at different levels, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. Upon receiving a legal request, cloud server will search the related indexes layer by layer instead of scanning all indexes.

V. 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 novel multi-keyword ranked search over the encrypted data based on hierarchical clustering index protocol (NMRSE-HCI).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 for storing all the data inside the database. The proposed application is divided mainly into following 3 modules. They are as follows:

1. Data Owner Module
2. Cloud Server Module
3. Data User/End User Module

1. DATA OWNER MODULE

This is the first module in which the data provider uploads their encrypted data in the Cloud server. For the security purpose the data owner encrypts the data file and then store in the server. The Data owner can have capable of manipulating the encrypted data file and performs the following operations Browse and enc and Uploads files, Grant Permission to cloud consumer / End user. Here the data owner has the privilege to upload the data in an encrypted manner into the cloud server and generate a trapdoor key which is of hexa decimal key for giving more security for the current data.

2. CLOUD SERVER MODULE

The Cloud server manages which is to provide data storage service for the Data Owners. Data owners encrypt their data files and store them in the Server for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the Server and then Server will decrypt them. The server will generate the aggregate key if the end user requests for file authorization to access and performs the following operations such as View all User Files, Give privileges to user, View Search Transaction, View all attackers, View all End Users, View all Data Owners, Create Index on searched data and provide all related data related to corresponding keyword, View all android users.

3. DATA USER/END USER MODULE

In this module we have a facility to access the data file with the secret key. The user can search the file for a specified keyword. The data which matches for a particular keyword will be indexed in the cloud server and then response to the end user. Here the data user can download the file from the cloud server if he/she got access privileges from the data owner and the cloud server. If the data user who has no access privilege from the data owner or cloud server, he/she cant able to access the data. Also the data user or end user can able to view the top K-documents that are available inside the cloud server based on hierarchical clustering approach.

VI. CONCLUSION

In this paper, we for the first time have implemented a novel multi-keyword ranked search over the encrypted data based on hierarchical clustering index protocol (NMRSE-HCI).This proposed mechanism is mainly used for online information retrieval as well as semantic search approaches.This current approach is also used for the correctness and completeness of search result.Also in our current thesis we implemented a encrypted search based on the keyword and also a ranked mechanism in which the data will be displayed in a ranked manner.This proposed protocol is efficient in identifying the top k-documents from the set of documents that are available in the cloud server in order to find out which document has the highest privilege and which document has the least privilege.By conducting various experiments on our proposed approach,we finally came to an conclusion that our proposed NMRSE-HCI approach is best in providing security for the cloud data in terms of data storage and data retrieval compared with security limitations that are available in current cloud servers.

VII. REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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).

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

[6] 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.

[7] "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.*

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

[9] 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.

[10] C. Wang, N. Cao, K. Ren, and W. J. Lou, Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467-1479,Aug. 2012.

[11] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, "Privacy- Preserving Multi-keyword Ranked Search over Encrypted Cloud Data," in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829-837.

[12] S. C. Yu, C. Wang, K. Ren, and W. J. Lou, "Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing," in Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1-9.

VIII. ABOUT THE AUTHORS

PALA SUPHALA is currently pursuing her 2 years M.Tech in Department of Computer Science and Engineering,D.N.R College of Engineering and Technology,Balusumudi, Bhimavaram, India. Her area of interest includes Cloud with Security.

D.D.D SURIBABU is currently working as an HOD & Assoc. Professor in Department of Computer Science and Engineering at D.N.R College of Engineering and Technology,Balusumudi, Bhimavaram, India. He has more than 14 years of teaching experience in engineering colleges.His research interest includes Data Mining, Big Data and Data Structures.