A Survey on the Architectures of Data Security in Cloud Storage Infrastructure

T.Brindha [#1], R.S.Shaji [#2], G.P.Rajesh [*3]
[#]Department of Information Technology, N.I.University, Thuckalay, TamilNadu, India.
[*]Department of Instrumentation and Control Engineering, Jayamatha Engineering College, Aralvaimozhi, TamilNadu, India.
[1]brindhahere@yahoo.com
[2]shajiswaram@gmail.com
[3]gprajesh84@gmail.com

***Abstract*— Cloud computing is a most alluring technology that facilitates conducive, on-demand network access based on the requirement of users with nominal effort on management and interaction among cloud providers. The cloud storage serves as a dependable platform for long term storage needs which enables the users to move the data to the cloud in a rapid and secure manner. It assists activities and government agencies considerably decrease their economic overhead of data organization, as they can store their data endorsement vaguely to third-party cloud storage sources than preserving the data centers on their personal. Cloud storage allows users to accumulate their data at a remote location and have the on-demand high superiority cloud requests without the burden of restricted hardware and software organization. Along with the widespread benefits on cloud computing, however, security issues with cloud data storage are arising in terms of confidentiality, integrity and reliability. Although the profits are obvious, such a service is also turning down user's corporeal control of their outsourced data, which unavoidably facades novel security dangers towards the accuracy of the data in cloud. This paper provides a comprehensive survey of major security issues that exists in cloud data storage. We also conduct extensive empirical studies and analysis of different techniques which provides security fortification for outsourced data in cloud computing environment.**

Keyword-Access control, Assured deletion, Backup/recovery, Distributed storage, Error localization.

I. INTRODUCTION

Cloud Computing is a web based application which provides computation, software, infrastructure, platform, devices and other resources to users on a pay-as-you-use basis. The cloud services can be utilized by the consumers without installation and their personal files can be access from any computer with internet access. The consumers are facilitated to use the services without installation and their individual files can be accessed from any computer with internet access. This technology provides more effective computing by organizing bandwidth and data storage and processing.

Among various kinds of cloud computing services, the dramatic growth of cloud storage services enables enterprises to outsource their data into cloud environment for rapid resource elasticity, location independent resource pooling and usage-based pricing. The data outsourced into the cloud offers more profits to consumers since they don't have to care about the adversities of direct hardware management. Even though, these cloud storage services facilitates large amount of space for storage and computing resources, it discards the responsibility for maintenance of data at the same time.

Hence for maintaining customer's data and information, the cloud storage systems are anticipated to meet various necessities such as scalability, high reliability, performance, replication, consistency, and data availability [12]. The security of data for such a storage service in cloud encompasses various aspects which includes secured channels, authority for access, and encryption. The cloud storage models stores the data on diverse virtualized servers.

The flow of data between the users and cloud servers and its storage in the cloud is shown in Fig. 1. The loss of control over security of data is problematic since any confidential and sensitive information provided by the consumer to the vendor will often be stored on, and accessible from, the vendor's systems. It's also necessary to assess the security features of a cloud storage service before making up a contract with the service provider. Big emphasis is placed by most organizations on securing the physical assets such as servers and storage, as well as securing their data assets from unauthorized access. Data security breaches if any, may lead to compliance and regulatory penalties. Hence it is therefore intensely necessary for the consumers to understand how data is being secured within the facilities provided by the cloud service provider.

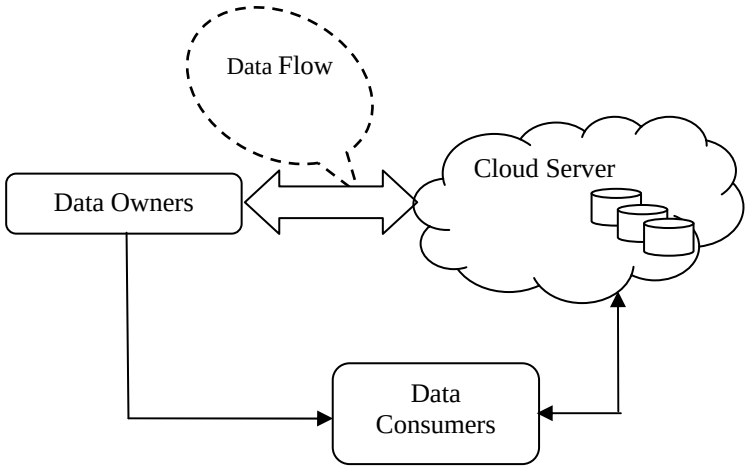

Fig. 1. Data storage over cloud

II. BACKGROUND WORKS

Cloud storage is a representation of systemized online storage where data is processed in virtualized pools of storage space which are usually hosted by third parties. Hosting corporations work on huge data centers, and people who need their data to be hosted, acquire or lease storage capability from them.

The cloud storage is mostly used under medicinal research records. Ming Li et al. [9] proposes an attribute based encryption technique to encrypt each person's health records which provides efficient revocation and access policy modification. Other techniques like key-policy attribute-based encryption (KP-ABE), ciphertext-policy ABE, ciphertext-policy attribute set-based encryption etc are also presented for flexible data access control.

In [14], Nallur et al. proposed a market-based mechanism is used in cloud-based service-oriented applications for self-adapting their QoS, depending on demand. The resource allocation problem is considered by Beaumont et al. in [11] that models both independent tasks scheduling and virtual machine allocation problems by introducing degree constraint. It proves that maintaining optimality is inexpensive. Recently, the significance of confirming the remote data integrity is pointed out using various security methods [2], [4], [6], [7], [10]. In general, the authentication mechanism only ensures the users for data access, nevertheless of assuring data security from the cloud vendor. Data security schemes are implemented [1], [3], [5] [8] [13] to secure data from the cloud vendor and other users of cloud.

The data stored in the clouds can be recovered without revealing the content to cloud service providers. In order to secure the data, Zhibin Zhou [17] proposed a privacy preserving cipher policy attribute-based encryption and an attribute based data storage (ABDS) system is presented that can be used for cryptographic access control mechanism. It minimizes cloud service charges by reducing communication overhead for data managements. The security challenges pretended by data-intensive applications deployed in cloud environments are explored by Wenchao Zhou et al. in [18].

The reported recent success of cloud computing has attracted attention for cost effective IT services. However, challenges are being faced by both research and professional communities including quality reliable services, optimized architectures and security. In [12], Cong Wang et al. highlight the security aspects of data storage from perspectives of threats and attacks from one side and approaches for solutions from the other side.

III. SECURITY CONCERNS ON CLOUD DATA STORAGE

Storage combines massive scalability and unparalleled flexibility with automated data placement, efficiently delivering content and information anywhere in the world through a standard interface. Cloud Storage eliminates the need to invest in expensive data storage infrastructure and associated maintenance costs. The result is the most flexible and cost efficient data storage solution on the market.

Security concerns on cloud storage may arise due to blowing of data growth and the demand of different data by the user. The different security challenges in cloud storage are:

- Snooping: Snooping is to intrude into others private data. It is a best way to send and retrieve the data over a secure transmission line.
- Data Leakage: Data leakage occurs when an uncontrolled and unauthorized data is transmitted between the user and the cloud. The best way is to encrypt the data at the starting point itself.

- Cloud Credentials: The encrypted data is also defenseless if it is combined with other's data. The clients can obtain others authorization and may delete the files. Hence it is necessary to protect one's unique authorization. The unauthorized clients must not be log in to others account and delete the data.
- Key Management: The cryptographic keys have to be managed in the cloud environment. The key management has to be very simple such that the generation and usage of keys should be automatic.
- Performance: An enduring security approach is necessary for encrypting and decrypting the files to and from the cloud but it must not affect the user's performance.

IV. ANALYSIS OF ARCHITETURES OF SECURE CLOUD STORAGE

As the services provided by cloud computing increases, the security concerns to the data in the cloud arises. The different architectures provided by different authors and the various approaches provided by them to protect the data in the cloud is presented here as follows:

A. Secure and Dependable Cloud Storage

In cloud environment, the data owners store their data in the cloud storage system. The storage correctness and data availability has to be ensured on the cloud servers.

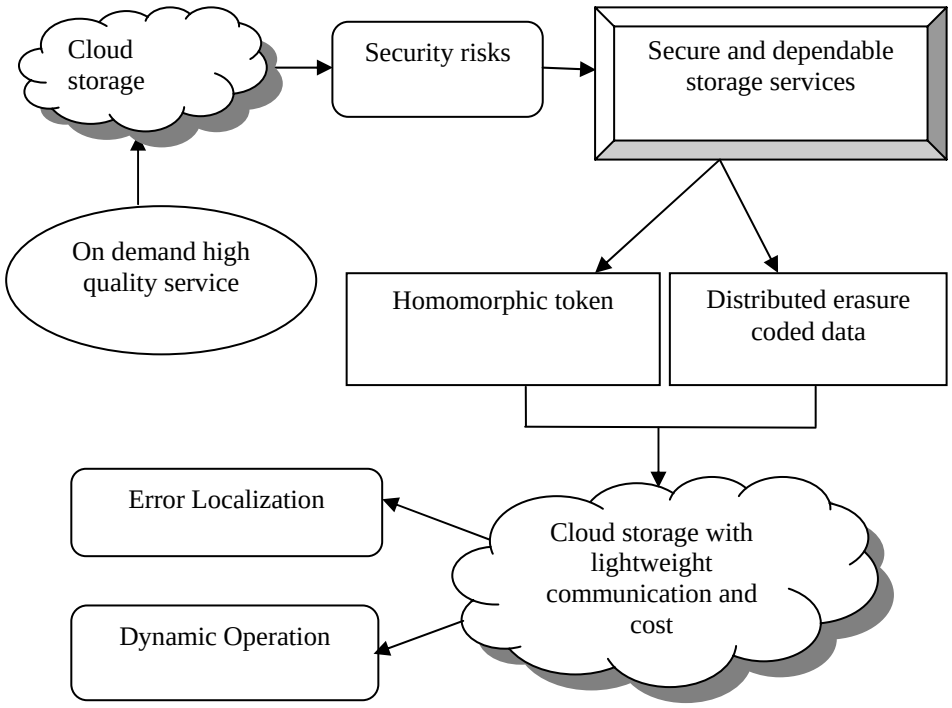

Fig. 2. Secure and dependable storage services

The most important problem arises with the detection of data updation by unauthorized users. To make certain the security and reliability for cloud data storage, competent mechanisms have been planned for active data verification [4] as shown in Fig. 2. For storing data in cloud, the Reed-Solomon erasure-correcting code approach is used to spread out the file F over a collection of dispersed servers. Small tokens are calculated for each distinct vector, V(i) where i Є (1, …, m). Based on the tokens, the unauthorized users can be identified if any inconsistencies are detected in the data storage, i.e., the tokens are compared with the obtained value to identify the unauthorized user.

Since the files are initially encrypted with a secret key k_s where i Є {1, 2, …, n} and then the Third Party Auditor (TPA) is assigned to perform auditing on the data files to check for the correctness of data. As the secret key is unknown to the TPA, there is no chance to the TPA for reading the content of the files during auditing. Moreover, the user has to be assured that all dynamics operation performed on cloud data storage has been processed by Cloud Service Provider (CSP) based on the storage verification tokens in a well-defined manner. The assurance of cloud data integrity and availability is achieved to increase the characteristics of cloud storage service.

B. Public Batch Auditing

Public auditing empowers the user as well as the external party to verify the integrity of data. But data privacy still remains a major issue. Cong Wang et.al [6] have presented an auditing protocol with privacy preserving approach which supports group auditing where different users can perform various auditing tasks at the same time. Each data files are encrypted with message authentication code (MAC) keys along with a secret key. The TPA receives the files with their MAC keys and audits the correctness of storage. The files are then verified for its integrity by using homomorphic linear authenticators based on a public key.

Instead of auditing the different files by different users individually for several times, multiple tasks are grouped together and batch auditing is performed. This approach hides the data blocks using random masking for ensuring the privacy of data to data owners. It also supports dynamic operations on cloud data including addition, modification and deletion. The batch auditing protocol uses the public key based homomorphic linear authentication in order to perform public auditing.

C. Integrity of Data in Multi-Cloud Storage

To store and maintain client's data, existence of multiple cloud service providers are considered in which a cooperative provable data possession (CPDP) is used to check the integrity and availability of stored data [21]. The data owner processes the file using a secret key and sends the file with some verification tags to the CSP. Then the CSP is requested to check the integrity of data with the help of a verification protocol. The procedure for verification is

1. The file is processed with a secret key to generate a set of procedure for verification and is transferred to the CSP.
2. By using the verification procedure, the files are audited for its integrity.

The CPDP scheme uses two algorithms namely key generation and tag generation [19] in which,

Key Generation: Provides a secret key k_s or a public secret key pair (k_p, k_s), k is the security parameter.

Tag Generation: A secret key k_s, a file S, and a set of cloud storage providers $S = \{S_k\}$ is taken as input and provides the triples (ζ, ψ, σ), where ζ is secret tags, ψ is the verification parameters and σ denotes set of all tags.

The protocol for data possession in cloud is represented in (1) as:

$$\left(\sum_{D_k \in D} D_k(I^{(K)}, \Omega^{(k)}) \leftrightarrow C\right)(k_p, \mu) = \begin{cases} 1 & I = \{I^{(K)}\} \text{ is intact} \\ 0 & I = \{I^{(K)}\} \text{ is changed} \end{cases} \quad \textbf{(1)}$$

where $I^{(k)}$ represents an input file, $\Omega^{(k)}$ indicates a collection of tags, D_k takes an input file and a tag as input, C results as true or false by indicating 1 or 0. k_p is the public key and μ is the input common for both D_k and C.

D. Fine Grained Access Control

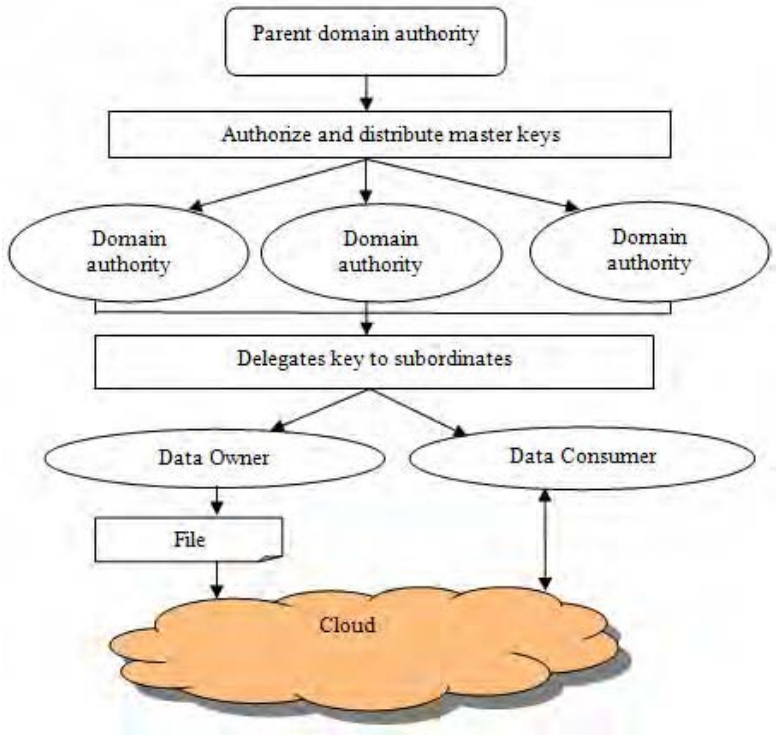

Fig. 3. Structure of system users

Access control is a strongly desired security requirement in cloud computing model. Wan et.al [19] proposed an encryption approach to handle the system users in a hierarchical manner as shown in Fig. 3. The system model consists of trusted authority, domain authorities, data owners and data consumers.

Initially the parent authority creates a public key p_k which is made public to the users and a master key m_k to be kept secret. The files are initially encrypted with data encryption keys (DEK) and then attribute based encryption is done on a hierarchical manner. To add a new user to the system, the administrating authority first check whether it is a valid one and then a unique ID is provided for the new user. When a request is received for data files from the user, the cipher texts for the data files are sent by the cloud which is decrypted to obtain DEK and then using DEK, the data files are decrypted. The data owner sends the file's unique ID and its signature to the cloud for deleting the encrypted data file. The cloud deletes the data file after successful verification of the data owner and the request.

E. Policy-based Access Control

To provide assurance for access control and assured deletion for outsourced data in cloud storage, the process of file assured deletion mechanism [20] is proposed by Yang Tang et al. The files are associated with file access policies which governs the files to be accessed. A data key is generated by the client which is used for encryption and decryption of data files. The control keys are maintained by key managers to encrypt or decrypt the data keys. The access keys are maintained by the client and are used to access the files of the associated policy. The key manager supports policy management in which policies, access keys and control keys are created or revoked and also performs key management in which the data keys are encrypted or decrypted.

For uploading the file, the client encrypts it according to the specified policy and then sends the file and the metadata to the cloud. These file and the metadata are retrieved for checking its integrity and then the file is decrypted for downloading it. This technique is also used for guaranteed deletion of files which cannot be possessed by any user even by the service providers. This scheme supports multiple policies in which the conjunction AND indicates that the data can be accessed if all the policies are satisfied and OR indicates data can be accessed if any policies are satisfied.

V. EVALUATION OF PARAMETERS

To perform the analysis of the schemes, the cloud is linked with various nodes. The auditing mechanism for integrity of the data storage [4] is analyzed based on security, efficiency and correctness.

- Data modification detection: The corruption of data can be highly detected by reducing the computational overhead on the server.
- Misbehaving server identification: If any unauthorized data modification occurs, the attack can be detected with high probability.
- Security strength: The storage correctness is ensured in a way that the cloud server stores the data in a proper way which will be audited by TPA. Moreover the privacy of data audited by TPA can be maintained since it cannot extract any information while auditing.

Group auditing [6] can be performed by preserving both storage correctness and privacy.

- Storage Correctness: It is proved that a reasonable reply cannot be provided by the cloud server to TPA unless the data are stored accurately.
- Privacy Maintenance: The original data of the user cannot be stemmed from the facts that are provided to the TPA for auditing.
- Security assurance for group auditing: The protocol used for group auditing ensures the same storage correctness and privacy maintenance for auditing multiple tasks together.

Flexible access policies can be used by the clients [21]. Scalability and efficient user revocation is provided for updating the user's key by adding a new expiration value to the existing key. Based on the optimization of block length and queries of probability, the performance of the integrity verification scheme [19] is analyzed. The communication and computation overhead can be greatly reduced and also reduces the extra storage size.

The simple pricing model is used to evaluate the financial overhead of assured deletion approach [20]. The time taken for performing cryptographic operations is increased when the size of the file increases. Based on file transmission time, metadata transmission time and cryptographic operation time, the time taken can be analyzed and cost can be evaluated using a backup system.

- Scalability: The task of generation of keys is distributed to low level authorities which reduces the workload of parent authority thus increasing the scalability.
- Flexibility: The access control scheme based on hierarchical approach combines the user attributes into a set and hence active bounds were imposed by the users which support multiple attributes conveniently.

- User revocation: An attribute of expiration time is just added to each key of the user and the authority maintains the user keys. The creation and dispersion of keys on a repeated basis can be prevented in this hierarchical approach.

VI. PERFORMANCE COMPARISON

The performance of the storage verification scheme [4] is accessed based on the cost of distribution of files as well as the generation of tokens. The average cost for computation of tokens by using homomorphic approach is nearly 0.4 ms. This technique is highly robust against unauthorized data modification attacks and byzantine failures.

Auditing multiple tasks at a time is tested based on time and the average time for auditing each task is calculated by dividing number of tasks from the total time for auditing. The time taken for auditing per task has been decreased to 15% in group auditing when compared to individual auditing at a time [6] as shown in Fig. 4.

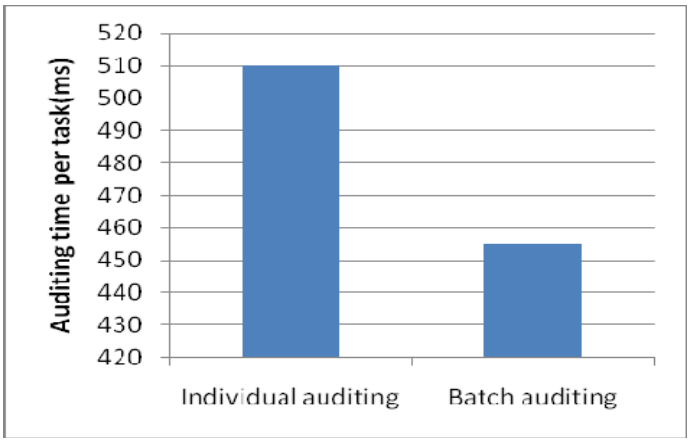

Fig. 4. Comparison of time taken for auditing

The feasibility of access control scheme [21] enhances the cloud storage security and also supports the organizations that outsources large amount of data to cloud. Based on the number of attributes and subsets, the cost is calculated and it has no impact on the access tree level. The computational complexity of different operations is shown in Table I.

TABLE I
Computation Complexity of Operations

Operation	Hierarchical Scheme
System setup	C(1)
Authority to old users	C(2X+Y)
Authority to new users	C(2X+Y)
Creation of files	C(2|Y|+|X|)
Revocation of users	C(1)
Deletion of files	C(1)

The results of the integrity verification scheme in [19] shows that storage and auditing performance have been increased, the computation overheads have been decreased and moreover the size of the extra storage space can also been reduced using the cooperative approach. The tampering data blocks can be easily identified by means of probabilistic verification scheme. The computation overhead of this scheme is reduced to nearly 3(n-1) times where n represents the number of clouds in a multicloud system.

The feasibility of policy based approach [20] contributes increased level of data security for existing cloud storage. It is appropriate for those organizations that need to archive enormous number of files with a limited amount of data. This approach is evaluated based on cost and overhead of running time. The transmission time is reduced if both the clients and the key manager are located within a region and overhead may increase if more details regarding the files are added to the metadata.

VII. DISCUSSIONS

In the existing work, security issues are handled for data operations, but not associated to the functional interpretation of the user data. Functional interpretation includes indirect data access through cloud applications, services rendering transaction, communication path, etc. In addition, data consistency in cloud data storage service environment is another major area of concern to improvise concurrent data application users. Data consistency state needs to be handled effectively for improving the low cost communication and computation users of the cloud.

The storage auditing mechanisms for integrity verification can be improved to a greater extent by enhancing the confidentiality and privacy of data that is being outsourced to the cloud. The policies used for access control can be enriched since the present schemes are less significant if the size of data file increases. Data integrity verification for large files still remains a challenging one.

VIII. CONCLUSION

In this paper, the different mechanisms presented by different authors are analyzed and compared that can be used in reducing the security risks in networked cloud storage. The problem of data security in cloud data storage which is basically a distributed storage system is summarized. Since the data can be outsourced, data backups off-site to intermediary cloud storage services decrease data management costs. Cloud data stores present scalability and high accessibility possessions for web submissions, but at the similar time they surrender data reliability. Nevertheless, the techniques we referred here cannot meet the expense of any data unpredictability. Data consistency state needs to be handled effectively for improving the low cost communication and computation users of the cloud. However, in the existing works, security issues are handled for data operations, but not associated to the functional interpretation of the user data.To meet the above challenges, the efficiency of secure and reliable cloud data storage framework can be improved for enhancing the concurrent data application users.

REFERENCES

[1] Bethencourt A. Sahai, and B. Waters, "Ciphertext-policy attribute based encryption," in *Proc. IEEE Symp. Security and Privacy*, Oakland, CA, 2007.

[2] Cong Wang, K. Ren, W. Lou, and J. Li, "Towards publicly auditable secure cloud data storage services," *IEEE Network Magazine*, vol. 24, no. 4, pp. 19–24, 2010.

[3] Cong Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for storage security in cloud computing," in *Proc. of IEEE INFOCOM'10*, San Diego, CA, USA, March 2010.

[4] Cong Wang, Qian Wang, Kui Ren, Ning Cao and Wenjing Lou, "Towards Secure and Dependable Storage Services in Cloud Computing", *Journal IEEE Transactions on Services Computing*, vol.5, pp.220-232, 2012.

[5] Cong Wang, Qian Wang, and Kui Ren and Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing", *Project Seminar on Cloud Computing*, Oct 2011.

[6] Cong Wang, S.M. Sherman, Chow, Qian Wang, Kui Ren and Wenjing Lou, "Privacy preserving public auditing for secure cloud storage" *IEEE transactions on Computers*, vol.62, pp.362-375, 2012.

[7] Deni Connor, Patrick H. Corrigan and James E. Bagley, "Cloud Storage: Adoption, Practice and Deployment", *An Outlook Report from Storage Strategies NOW*, April -4, 2011.

[8] V. Goyal , O. Pandey, A.Sahai, and B.Waters, "Attribute-based encryption for fine-grained access control of encrypted data," in *Proc. ACM Conf. Computer and Communications Security (ACM CCS)*, Alexandria, VA, 2006.

[9] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou, "Scalable and Secure Sharing of Personal Health Records in Cloud Computing using Attribute-based Encryption", *IEEE Transactions on Parallel and Distributed Systems*, 2012.

[10] Nashaat el-Khameesy, Hossam Abdel Rahman, "A Proposed Model for Enhancing Data Storage Security in Cloud Computing Systems", *Journal of Emerging Trends in Computing and Information Sciences*, Vol 3, No. 6, June 2012.

[11] Olivier Beaumont, Lionel Eyraud-Dubois and Hejer Rejeb, "Heterogeneous Resource Allocation under Degree Constraints", *IEEE Transactions on Parallel and Distributed Systems*, 2012.

[12] G.A Patil, S.B Patil, "Data Security Mechanism for Cloud", International Conference on Emerging Technology Trends (ICETT), *Proceedings published by International Journal of Computer Applications (IJCA)*, 2011.

[13] Qian Wang; Cong Wang; Kui Ren; Wenjing Lou; Jin Li "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing", *IEEE Transactions on Parallel and Distributed Systems*, Vol 22, Issue 5, 2011.

[14] Vivek Nallur, Rami Bahsoon, "A Decentralized Self-Adaptation Mechanism for Service-Based Applications in the Cloud", *IEEE Transactions on Software Engineering*, 2012.

[15] Wenchao Zhou, William R. Marczaky Tao, "Towards Secure Cloud Data Management", *Department of Computer & Information Science, Technical Reports (CIS)*, University of Pennsylvania, 2010.

[16] K.Yang, X. Jia, "An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing", *IEEE Transactions on Parallel and Distributed Systems*, Vol. pp, Issue. 99, 2012.

[17] Yan Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S.S. Yau, "Dynamic Audit Services for Integrity Verification of Outsourced Storages in Clouds," *Proc. ACM Symp. Applied Computing*, pp. 1550-1557, 2011.

[18] Yan Zhu, H. Hu, G.J. Ahn, Y. Han, and S. Chen, "Collaborative Integrity Verification in Hybrid Clouds," *Proc. IEEE Conf. Seventh Int'l Conf. Collaborative Computing: Networking, Applications and Work sharing*, pp. 197-206, 2011.

[19] Yan Zhu, Hong Xin Hu, Gail-Joon Ahn and Meng Yang Yu, "Cooperative Provable Data Possession for Integrity Verification in Multi cloud Storage", *IEEE Transactions on Parallel and Distributed System*, Vol 23, pp-2231-2244, 2012.

[20] Yang Tang, Patrick P.C. Lee, John C.S. Lui, Radia Perlman "Secure Overlay Cloud Storage with Access Control and Assured Deletion", *IEEE Transactions on Dependable and Secure Computing*, vol.9, pp. 903-916, 2012.

[21] Zhiguo Wan, Jun'e liu and Robert H.Deng, "HASBE: A hierarchical attribute-based solution for flexible and scalable access control in cloud computing", *IEEE transactions on information forensics and security*, vol.7, no.2, April 2012, pp.743-754.

[22] Zhibin Zhou and Dijiang Huang, "Efficient and Secure Data Storage Operations for Mobile Cloud Computing", *IACR Cryptology ePrint Archive*, 2011.