International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
Article Received: *02 January 2023* ***Revised:*** *12 February 2023* ***Accepted:*** *02 March 2023*

Enhancing Privacy in Big Data through an Asymmetric Secure Storage Protocol with Data Sharing

Kanigiri Suresh[1], Dr. Manoj Eknath Patil [2]
Department of Computer Science & Engineering, Dr. A.P.J. Abdul Kalam University, Indore (M.P.) - 452010

Abstract: Cloud computing has become integral to handling large-scale data in the era of big data. Storing such vast amounts of data locally is cost-prohibitive, necessitating the use of cloud storage services. However, reliance on a single cloud storage provider (CSP) raises concerns such as service interruptions and security vulnerabilities, including insider threats. To address these issues, this research proposes a novel approach where big data files are distributed across multiple CSPs using an asymmetric security framework. Metadata encryption and decentralized file access management are facilitated through a dew computing intermediary, enhancing security and ensuring privacy. Unlike previous approaches, this protocol employs group key encryption and secret sharing schemes within SSGK for efficient data protection and access control. Extensive security and performance evaluations demonstrate significant reductions in security risks and privacy breaches while optimizing storage efficiency.

Keywords*: Big Data, Cloud Computing, Cloud Security, Cloud Storage Provider (CSP), Asymmetric Security*

I. INTRODUCTION

Big data represents an extensive and continuously growing collection of data characterized by its massive volume and diverse formats, including structured, semi-structured, and unstructured data types [2]. As the volume of data generated daily increases exponentially, traditional database systems struggle to handle such immense datasets efficiently and cost-effectively, necessitating alternative storage solutions [1]. Cloud computing emerges as a viable option by providing scalable and cost-efficient storage services capable of accommodating big data requirements across various applications and industries [4].

Cloud computing, alongside technologies like Business Intelligence, Data Mining, Industrial Information Integration Engineering (IIIE), and Internet-of-Things (IoT), has ushered in a new era for Enterprise Systems (ES) by offering dynamic resource allocation and exceptional scalability compared to traditional distributed systems [3]. Cloud-based solutions enable distributed query processing and rapid data retrieval, leveraging virtualized environments to store big data across cloud storage providers (CSPs) rather than locally on individual devices [5]. This approach not only optimizes storage costs but also enhances data accessibility and processing efficiency through distributed storage technologies.

Despite these advantages, storing sensitive data in cloud environments raises significant security and privacy concerns [7]. Issues such as vendor lock-in, where reliance on a single CSP for data storage can lead to service disruptions or increased costs due to provider policy changes, highlight the need for robust security measures [6]. Cloud storage providers implement encryption and security protocols to protect data from external threats, yet challenges remain regarding insider threats and data integrity within CSPs [8]. These challenges necessitate solutions that ensure data privacy, access control, and secure data sharing practices across multiple cloud platforms.

In this research, we propose a multi-cloud approach to mitigate vendor lock-in risks and enhance data security and access control for big data storage [10]. Data files are fragmented into smaller chunks and distributed across multiple CSPs, with metadata encrypted to secure information about chunk locations and access paths [9]. Utilizing asymmetric encryption techniques, metadata encryption ensures efficient key management and access authorization without the overhead of encrypting entire data files, thereby optimizing data retrieval and security measures [11]. Moreover, secret sharing schemes further safeguard data by distributing decryption keys among authorized users, preventing unauthorized access and enhancing overall data confidentiality on cloud storage platforms.

By integrating these advanced security mechanisms into cloud storage services, this protocol aims to establish a security-aware cloud environment capable of meeting

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 02 January 2023 **Revised:** 12 February 2023 **Accepted:** 02 March 2023*

stringent data privacy requirements for mission-critical business applications [12]. This approach not only addresses current security challenges in cloud computing but also paves the way for secure and efficient data management practices in the era of big data analytics and cloud-based services

II. LITERATURE SURVEY

K. D. Bowers, A. Juels, and A. Oprea et.al [16] descried HAIL system that upon high availability and integrity protection within the cloud. Also, data privacy is not of primary concern. It is a distributed cryptographic system that allows a set of servers to show a client that a stored file is unimpaired and retrievable. Data is distributed and split by using erasure codes, similar to the method in RACS, upon multiple clouds to achieve high availability. Data stored on a single server is also redundantly stored to increase its resistance against bitrot. A proof of retrievability protocol based on active servers and proofs of data possession has been developed to confirm the availability and correctness of data,

Y. Singh, F. Kandah, and W. Zhang et.al [15] analyzed an economic data distribution model among the available CSPs in the market. This scheme provides customers with data availability and secure storage. In SCMCS model, the customer divides and distributes his or her data among several CSPs available in the market. However, data owner have to consider CSP selection based on his available budget. SCMCS provides a decision for the customer which CSPs can be selected concerning data access quality of service offered by the CSPs at the location of data retrieval. This not only rules out the likelihood of a CSP misusing the customers' data, breaching the privacy of data but can efficiently ensure the data availability with a better quality of service.

R. Pottier and J. M. Menaud et.al [4] described Trusty Drive model is a storage system based on many cloud providers to provide users with data privacy as well as reliable storage. In this architecture, the data privacy is determined by two rules: the document anonymity and the user anonymity. The document anonymity guarantees that the storage system and cloud providers do not know about stored documents, and content inside the documents. The user anonymity protects users against linking users and stored documents. To achieve this anonymity, users divide their documents among several cloud providers to yield that no provider hosts the whole document. The documents splitting is completed at the user level. The storage system is not able to reconstruct user documents. To increase the document anonymity, the system lets users choose the encoding process of their data in order to deal with illegible blocks.

V. R. Balasaraswathi and S. Manikandan et.al [8] presented the cryptographic data splitting with dynamic approach for securing information in hybrid cloud. The application data is partitioned and distributed to distinct clouds, which is the public cloud. Data can only be partitioned using classical cryptographic methods, AES. AES encrypt the user file with the key length of 256 bits and then sliced encrypted into pieces. A private cloud holds the metadata information. The metadata information is passwords, secret keys of each file, and encrypted access paths reside in private cloud securely. This approach prevents the unauthorized data retrieval by hackers and intruders.

D. Sánchez and M. Batet et.al [5] introduced a semantically- grounded data splitting mechanism that can automatically detect pieces of data that may cause privacy risks and split them on local premises. Chunks of clear data are independently stored into the separate locations of a multi- cloud. External entities cannot access to the whole confidential data. This scheme is applied to medical record, which requires high level of privacy.

K.Huang, R.Tso, Y.Chen, et al [9] introduced a novel public key encryption with authorized equality warrants on all of its ciphertext or a specified ciphertext. Which strengthen the securing requirement.

K.Xue, Y.Xue, W.Li, et al [6] proposed a new framework, named RAAC, to eliminate the sigle-point performance bottleneck of the exiting CPABE based access control schemes for public cloud storage. While these schemes use identity privacy by using attribute based techniques which fail to protect user attribute privacy.

Pervez Z., Khattak A. M. and Lee S et.al [14] addressed the privacy issues in a cloud-based storage a privacy aware data sharing scheme SAPDS. It combines the attribute based encryption along with proxy reencryption and secret key updating capability without relying on any trusted third party. But the storage and communication overhead of SAPDS is decided by attribute encryption scheme.

Jeong-Min Do, You-Jin Song , Namje Park et, al [12] proposed the data confidentiality solution using proxy re-encryption approach. Which allows transitive encryption means any number of time encryption and decryption is performed.

J.Shao, R.Lu and X.Lin et.al [10] proposed Light weight encryption approach to cloud environment. Their approach consumes less power because of key size is very less.

However it does not comply with data security and privacy needs of cloud.

A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, et.al [13] analyzes an object-store interface on top of passive storage clouds. Its data objects utilize cryptographic hashes for integrity control; short- time version numbers provide for concurrent updates. System model uses an asynchronous distributed system composed of three types of parties: writers, readers, and cloud storage providers. As no active server components can be used, the system cannot cope with malicious writers. Multiple concurrent writers are supported through client-side locks: this allows for obstruction- free, but not wait- free, operation. Cloud providers are allowed to fail in Byzantine ways. Confidentiality is optionally supported by secret- sharing techniques in the DepSky-CA variant.

Table-1 Summarized Literature Review

Authors	Key Findings	Research Gap
K. D. Bowers, A. Juels, A. Oprea, et al.	Developed HAIL system for cloud integrity using distributed cryptographic systems and proof of retrievability protocols.	Focuses on integrity and availability, but less emphasis on data privacy.
Y. Singh, F. Kandah, W. Zhang, et al.	SCMCS model for economic data distribution among CSPs, ensuring data availability and secure storage.	Lack of consideration for dynamic CSP selection based on changing service quality and user requirements.
R. Pottier, J. M. Menaud, et al.	Trusty Drive model ensuring document and user anonymity by distributing data across multiple cloud providers to enhance privacy.	Limited exploration of scalability issues with increasing number of cloud providers and impact on performance.
V. R. Balasaraswathi, S. Manikandan, et al.	Cryptographic data splitting using AES for data security in hybrid clouds, with metadata stored securely in private clouds.	Does not address the overhead and latency introduced by splitting and managing metadata in hybrid cloud environments.
D. Sánchez, M. Batet, et al.	Semantically-grounded data splitting for privacy-sensitive data, particularly in medical records, ensuring distributed storage across multiple clouds.	Lacks evaluation on scalability and performance implications when applied to large-scale datasets and real-time data access scenarios.
K. Huang, R. Tso, Y. Chen, et al.	Introduces public key encryption with authorized equality warrants, enhancing security guarantees for ciphertexts in cloud storage.	Limited exploration on practical implementation challenges and performance impacts on cloud storage operations.
K. Xue, Y. Xue, W. Li, et al.	RAAC framework for enhancing access control efficiency in public cloud storage, addressing privacy concerns in attribute-based access control schemes.	Insufficient evaluation on scalability and compatibility with emerging cloud architectures and distributed storage systems.
Pervez Z., A. M. Khattak, S. Lee, et al.	SAPDS scheme combining attribute-based encryption, proxy re-encryption, and key updating for secure and privacy-aware data sharing in cloud storage.	Needs further exploration on scalability under varying workload conditions and integration with diverse cloud service models.
Jeong-Min Do, Y. J. Song, N. Park, et al.	Proxy re-encryption for enhancing data confidentiality in cloud environments, allowing for transitive encryption and decryption.	Limited analysis on the overhead and performance implications of frequent re-encryption operations in dynamic cloud environments.
J. Shao, R. Lu, X. Lin, et al.	Lightweight encryption approach for reducing power consumption in cloud environments, addressing performance issues with smaller key sizes.	Requires further investigation into the trade-offs between lightweight encryption and robust data security measures in cloud storage.
A. Bessani, M. Correia, B. Quaresma, et al.	DepSky-CA variant using object-store interface and secret-sharing for data confidentiality in passive storage clouds, accommodating Byzantine failures.	Insufficient exploration on the impact of Byzantine failures on system performance and scalability in large-scale cloud deployments.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 02 January 2023 **Revised:** 12 February 2023 **Accepted:** 02 March 2023*

Research Gap Analysis:

1. **Privacy Emphasis**: Many existing solutions focus primarily on data integrity and availability in cloud storage, often overlooking comprehensive privacy protections, especially in scenarios involving dynamic data sharing and access.
2. **Scalability Challenges**: Few studies comprehensively address scalability issues with increasing numbers of cloud providers or the impact on performance and latency in real-time data access scenarios.
3. **Dynamic CSP Selection**: There is a need for more research into dynamic CSP selection mechanisms based on changing service quality, cost considerations, and user-specific requirements for data availability and privacy.
4. **Performance Optimization**: Further exploration is needed to optimize performance overheads introduced by cryptographic operations, metadata management, and data splitting techniques in hybrid and multi-cloud environments.
5. **Integration Challenges**: Limited exploration on the integration challenges of existing cryptographic techniques with emerging cloud architectures and distributed storage systems.

   These gaps highlight opportunities for future research to enhance the security, privacy, scalability, and performance of cloud storage solutions, particularly in the context of big data and sensitive information handling.

III. Asymmetric Secure Storage Scheme Using Data Sharing Protocol For Privacy Risks In Big Data.

In Asymmetric Secure Storage Scheme using Data Sharing Protocol for Privacy Risks in Big Data, fig.1 file distribution on multi- cloud is represented and in fig.2 cloud storage for big data is represented.

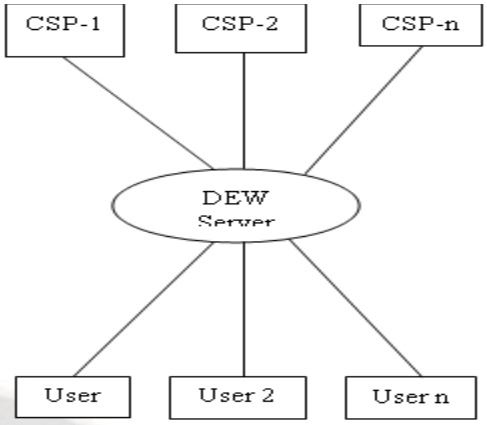

Fig.1 Block Diagram Of File Distribution On Multi- Cloud

Dew server is light weight server that monitor availability of data chunks, manage request access from users, send authorize token from data owner to request users.

Cloud storage provider offer storage service for customers. Different cloud service providers have different policies, promotions and operating cost. Cloud storage provider provide application interface (API) for developers to connect their application to its service.

User is a person who request to access the sharing file on multi cloud storage. The user must have authorization from data owner before access to the data chunks. After he get authorization, he can use information in authorize token to connect to cloud storage providers that keep data chunks.

Firstly, data distribution, data owner opens *n* network connections to cloud storage providers, then upload pieces of data into different CSPs. Secondly, recollect steps, after user receive authorize token from dew server, user create *n* network connection to cloud storage providers to retrieve *n* pieces of data chunk. This retrieval is performed concurrently. The public parts are stored in multiple cloud storages. The secret part is stored locally and protected by data owner. The public part contains the content of the file such as text, database, pictures, voice or video. Our scheme will distribute the public part into several cloud storages. Each cloud storage provider holds a part of the public data part.

The architecture is hybrid between client/ server architecture and peer-to-peer architecture. There is host act like server and clients as in client/ server scheme. Also, there is n CSPs act like seeder and user clients act like leecher in peer-topeer scheme. However, there is some different characteristic. This architecture has no contribution data transfer among seeders and leechers.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3
***Article Received:** 02 January 2023 **Revised:** 12 February 2023 **Accepted:** 02 March 2023*

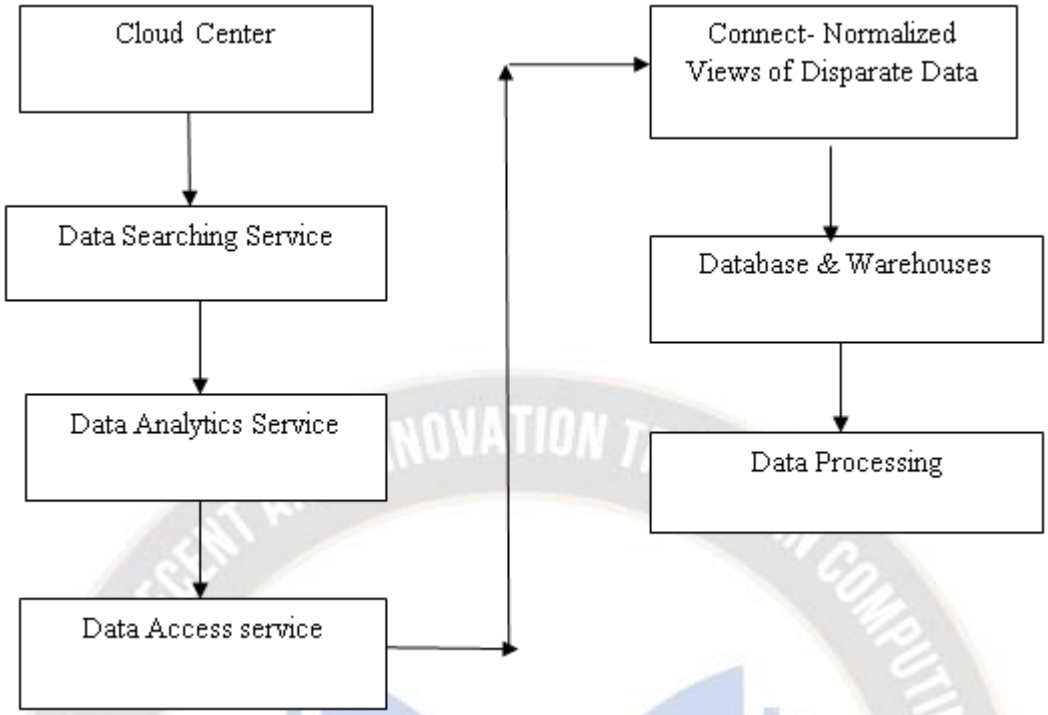

Fig.1: Block Diagram Of Cloud Storage For Big Data

According to the fig.2 the cloud center receives different kind of data from the multiple protocol formats. The next data searching service is done. Structure is any data structure that allows the efficient
retrieval of specific items from a set of items, such as a specific record from a database. The simplest, most general, and least efficient search structure is merely an unordered sequential list of all the items.

Data analytics is the process of examining datasets to draw conclusions about the information they contain. A data analytic technique enables to take raw data and uncover patterns to extract valuable insights from it. Data access refers to a user's ability to access or retrieve data stored within a database or other repository. Users who have data access can store, retrieve, move or manipulate stored data, which can be stored on a wide range of hard drives and external devices. Data Access Service (DAS) simplifies handling of data when interacting with the back-end data source and frees application developers from dealing with tedious and error-prone transformation between end source types and SDO Data Object Types/properties.

Connect normalized views of disparate data, which means the data normalization is a method in which data attributes are structured to improve the cohesion of the types of entities within a data model. In other words, the purpose of data standardization is to minimize and even eradicate data duplication, an important factor for application developers because it is extremely difficult to store items in a relational database that contains the same data in many locations.

A database and warehouse is any collection of data organized for storage, accessibility, and retrieval. A data warehouse is a type of database the integrates copies of transaction data from disparate source systems and provisions them for analytical use.

Data processing is a way to manage data across cloud platforms, either with or instead of on-premises storage. The cloud is useful as a data storage tier for disaster recovery, backup and long-term archiving. With cloud data management, resources can be purchased as needed.

IV. RESULT ANALYSIS

The result analysis of framework of Asymmetric Secure Storage Scheme using Data Sharing Protocol for Privacy Risks in Big Data is demonstrated in this section.

Table.1: Performance Analysis

Classifiers	Security	Generation Time (Sec)

Data Sharing Protocol for Privacy Risks in Big Data	98	14
ACPC for Privacy Risks in Big Data	91	25

The above table shows that the performance analysis of the Asymmetric Secure Storage Scheme using Data Sharing Protocol for Privacy Risks in Big Data gives high security and less generation time.

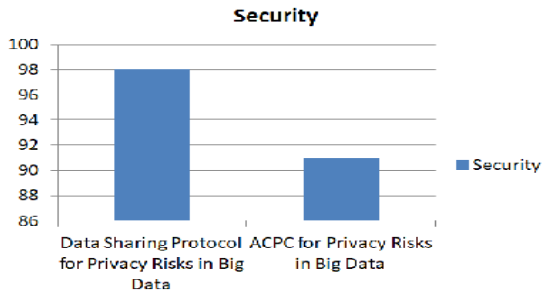

Fig.2: Security Comparison Graph

In Fig.2 security comparison graph the security for a framework of on Asymmetric Secure Storage Scheme using Data Sharing Protocol for Privacy Risks in Big Data shows higher security.

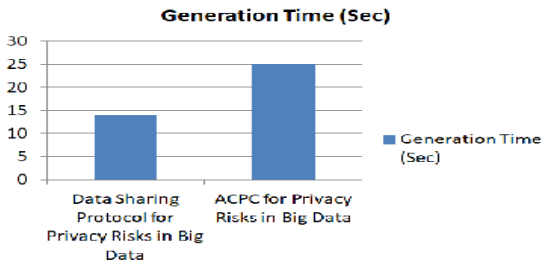

Fig.3: Generation Time Comparison Graph

In this comparison the above graph shows that the framework of Asymmetric Secure Storage Scheme using Data Sharing Protocol for Privacy Risks in Big Data shows low generation time when compared with other methods.

V. CONCLUSION

In this section, we conclude on our asymmetric secure storage scheme utilizing a data sharing protocol to mitigate privacy risks in Big Data. Our approach involves distributing Big Data files across multiple cloud storages by dividing them into equal chunks. We conducted a comprehensive analysis of both security and performance aspects. Our system achieves robust security with minimal complexity, ensuring that data privacy is maintained effectively. Compared to traditional client/server architectures, our scheme demonstrates superior performance in terms of both storage efficiency and computational overhead. Encryption secures data transmission over public channels, while our verified security mechanisms ensure that only authorized parties can access the data grids stored in cloud storage. The efficiency gains in storage and computation underscore the practicality and effectiveness of our approach, resulting in high security standards and reduced data generation times.

REFERENCES

[1] F. R. Damayanti, K. A. Elmizan, Y. F. Alfredo, Z. N. Agam and A. Wibowo, "Big Data Security Approach in Cloud: Review," *2018 International Conference on Information Management and Technology (ICIMTech), Jakarta, 2018*,pp.428-431.

[2] V. C. Storey and I.-Y. Song, “Big data technologies and Management: What conceptual modeling can do,” *Data & Knowledge Engineering*, vol. 108, pp.50-67, March 2017.

[3] Peng Zhao, Wei Yu, Shusen Yang and Xinyu Yang, Jie Lin, "On Minimizing Energy Cost in Internet-Scale Systems With Dynamic Data," Access IEEE., vol. 5, pp. 20068-20082, 2017.

[4] R. Pottier and J. M. Menaud, “TrustyDrive, a multi-cloud storage service that protects your privacy,” *IEEE Int. Conf. Cloud Comput. CLOUD*, pp. 937–940, 2017.

[5] D. Sánchez and M. Batet, “Privacy-preserving data outsourcing in the cloud via semantic data splitting,” *Comput. Commun.*, vol. 110, pp. 187– 201, 2017.

[6] K.Xue, Y.Xue, W.Li,.,"RAAC:Robust and Auditable Access Control with Mutiple Attrbute Authoroties for Public Cloud Storage," IEEE trans. on Info. Forensics and Security, vol.12, no.4, april,2017.

[7] Z.Fu, K.Ren , J.Shu, X.Sun, et al,"Towards efficient content-aware search over encrypted outsourced data in cloud," in Proc.IEEE Conf. Comput. Commun (INFOCOM),Apr.2016,pp.1-9.

[8] V. R. Balasaraswathi and S. Manikandan, “Enhanced security for multicloud storage using cryptographic data splitting with dynamic approach,” *Proc. 2014 IEEE Int.*

Conf. Adv. Commun. Control Comput. Technol. ICACCCT 2014, no. 978, pp. 1190–1194, 2015.

[9] K.Huang, R.Tso, Y.Chen, "PKE-AET: public key encryption with authorized equality test," The computer Journal, vol.58, no.10, pp.2686- 267,2015.

[10] J.Shao, R.Lu and X.Lin, "Fine-grained data sharing in cloud computing for mobile devices," in Proc IEEE Conf, Comput. Commun. (INFOCOM), Apr 2015, pp.2677-2685.

[11] A. Kanai, N. Kikuchi, S. S. Tanimoto, and H. Sato, "Data Management Approach for Multiple Clouds Using Secret Sharing Scheme," *2014 17th Int. Conf. Network-Based Inf. Syst.*, pp. 432–437, 2014.

[12] Jeong-Min Do, You-Jin Song , Namje Park ,"Attribute based Proxy Re-Encryption for Data Confidentiality in Cloud Computing Environment" ,first ACIS/JNU International Conference on Computers, Networks, Systems, and Industrial Engineering.

[13] A. Bessani, M. Correia, B. Quaresma, F. André, and P. Sousa, "DepSky:Dependable and Secure Storage in a Cloud-of-Clouds," *ACM Trans. Storage*, vol. 9, no. 4, pp. 1–33, 2013.

[14] Pervez Z., Khattak A. M. and Lee S., "SAPDS: self-healing attribute-based privacy aware data sharing in cloud," The Journal of Supercomputing, vol.62, no.1, pp.431-460, Oct.2012.

[15] Y. Singh, F. Kandah, and W. Zhang, "A secured cost-effective multicloud storage in cloud computing," *2011 IEEE Conf. Comput. Commun. Work. INFOCOM WKSHPS 2011*, pp. 619–624, 2011.

[16] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," *Proc. 16th ACM Conf. Comput. Commun. Secur. - CCS '09*, vol. 489, p. 187, 2009.

[17] R. Kumar and K. W. Ross, "Optimal peer- assisted file distribution: Single and Multi-class problems," *Proc. IEEE Work. Hot Top. Web Syst. Technol.*, pp. 1–11, 2006.

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

[19] Sahai A,Waters B. Fuzzy identity-based encryption. Proceeding of EUROCRYPT 2005. Berlin : Springer 2005,LNCS 3494:457-473.