JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 07, ISSUE 19, 2020

ASYMMETRIC SECURE STORAGE SCHEME USING DATA SHARING PROTOCOL FOR PRIVACY RISKS IN BIG DATA.

Research Scholar - 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: Recently, cloud computing is an emerging technology along with big data. Both technologies come together. Due to the enormous size of data in big data, it is impossible to store them in local storage. Alternatively, even we want to store them locally, we have to spend much money to create bit data center. One way to save money is store big data in cloud storage service. Cloud storage service provides users space and security to store the file. However, relying on single cloud storage may cause trouble for the customer. CSP may stop its service anytime. Also, the CSP is the third party that user have to trust without verification. After deploying his file to CSP, the user does not know who access his file. Even CSP provides a security mechanism to prevent outsider attack. However, how user ensure that there is no insider attack to steal or corrupt the file. The big data file is split into chunks and distributed to multiple cloud storage provider (CSP). Asymmetric security concept is applied to this research. The metadata will be encrypted and transfer to the user who requests to access the file. The file accessing, monitoring, metadata transferring is functions of dew computing which is an intermediate server between the users and cloud service. Different from the prior works, group key is used to encrypt the shared data and secret sharing scheme is used to distribute the group key in SSGK. Extensive security and performance analyses indicate that our protocol highly minimizes the security and privacy risks of sharing data in cloud storage and saves the storage space. Hence, the analyzes minimize the risk; ensure data privacy, also accessing control and less generation time.

KEYWORDS: Big Data, Cloud Computing, Cloud Security, cloud storage provider (CSP)

I. INTRODUCTION

Big data is a known asset of data that is massive in size and with the passage of time yet growing exponentially. Big Data could be Structured, Semi-structured, and Unstructured. Millions of data are uploaded on a daily basis. An Immense amount of data is sent and received. In short, big data gives the idea about when we have a large amount of data in a form of structured, Semistructured and unstructured, due to its large volume and complex nature of data, It is very difficult to handle a massive amount of data and this task is not performed by using traditional database system [2]. A large server is required for the implementation of big data which needed high maintenance and cost. The cloud server can become a potential choice to solve this problem [1].

The emerging technologies about big data such as Cloud Computing, Business Intelligence, Data Mining, Industrial Information Integration Engineering(IIIE) [4] and Internet-of-Things have opened a new era for future Enterprise Systems(ES) [3]. Cloud computing is a new computing model, in which all resource on Internet form a cloud resource pool and can be allocated to different applications and services dynamically. Compared with traditional distribute system, a considerable amount of investment saved and it brings exceptional elasticity, scalability and efficiency for task execution. By utilizing Cloud Computing services, the numerous enterprise investments in building and maintaining a supercomputing or grid computing environment for smart applications can be effectively reduced.

Cloud computing and big data are associated. Big data facilitate users by providing the ability to use specialty computing to process distributed queries across multiple set of data and return resultant assortments promptly. Big data utilizes distributed storage technology based on cloud computing rather than local storage attached to a computer or electronic device. Big data evaluation is driven by fast- growing cloud- based applications developed using virtualized technologies. A way to store big data is depositing data in cloud storage instead of spending substantial funds on building a massive data center in the organization. Cloud storage is a crucial application of cloud computing. Consumers can store their files on cloud storage provider (CSP). CSP implements storage- as- a service which supports database technologies, both SQL and NoSQL. After omitting their files on CSP, data owners deliver the maintaining storage duty to the CSP.

Despite these advantages, security requirements dramatically rise when storing personal identifiable on cloud environment. This raise regulatory compliance issues since migrate the sensitive data from federate domain to distribute domain. To take the benefit enabled by big data technologies, security and privacy issues [7] must be addressed firstly.First concern is lock-in vendor. Lock-in vendor is a situation that consumer relies on and host all his or her files in one specific CSP. CSP persuades the customer to use its services by promotions or benefits. However, CSP may change their services or agreement such as, prices, bandwidth, service time. Data loss might occur in CSP data center due to disaster or war. Technology threats could harm CSP in many forms such as cyber-attack, hacking. Financial of CSP is also a significant issue. These cause CSP stops storage service anytime. The mentioned issues would affect service usage of the customer. The customer might lose data or face difficulty use of services. It is hard and costly for moving a large of files from one CSP to another CSP. The customer has to pay transfer charge from former CSP (outbound) to brand-new CSP (inbound). It means that customer necessity pays two times or more to migrate all files.

The second risk issue of hosting file on one CSP is trustworthy. Even the CSP guarantee data security for the customer by implement security or safety technologies. Cloud storage providers often offer a general encryption of all data stored on their servers using a company key which is known only to them. This may prevent data stealing from external attackers but does not protect against any attacks which include theft of the encryption key or internal attacks conducted by personnel who are able to gain access to these keys. Cloud storage service is only the third parties provide storage services for the user. The user has to place data into the provider whose reliability is hidden. Consequently, when there are malicious actions inside the storage provider, a user's data can be damaged or corrupted easily. These actions happen inside without notice of user. It is common that management and operation are not exposed, which means the user cannot verify them. Therefore, a user must trust the declaration of a provider without attestation.

The lack of trustworthy in CSP leads to privacy and copyright concern issues. Due to unexposed operations inside CSP, how can data owner ensure that only authorized users to access his or her data? If data stored on a CSP is confidential data, only granted users can read, copy or even modify it. Data owner can control file accessibility if it is kept in local storage. In the situation that data is stored remotely on cloud storage, the data owner cannot know or detect file accessibility form inside storage provider [11].

In this research, we propose storing big data on multi- cloud to overcome vendor lock- in, data security, also access control. Data owner separates file to chunks then distribute them on multiple cloud storages. The metadata file or secret part is generated and kept secretly. The asymmetric security thought is applied in this study. Alternatively, performing encryption on the data file is an inappropriate way. Due to the enormous volume or size of big data, it is almost impossible to encrypt the entire file. Even though, we can encrypt the big file. It is time consuming for the user to decrypt the encrypted file back to original form. Instead of doing this, we can encrypt only the metadata file, which contains the location of chunks, access paths, and other related information. The size of metadata comparing with the whole file is significantly small. So, it is more practical to encrypt the metadata file rather than the big data file. Metadata contain the location of each chunk and access paths. If a user requires accessing the file, data owner will send metadata manually or automatically to the user. The user uses metadata as a key and map to grant him access to chunks on multiple cloud storage and retrieve chunks into his machine.

Asymmetric encryption algorithms are used to encrypt the interactive message and makes only legitimate participants have the ability to decrypt the key. Thirdly, in case of shared data being known by unauthorized users, this protocol uses secret sharing scheme to assign key to the legitimate participants. By adding security mechanism to conventional service oriented clouds, we obtain a security aware cloud and guarantee the privacy of data sharing on cloud storage. Building security mechanism on cloud storage may accelerate the deployment of a cloud in mission critical business scenario.

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.

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.

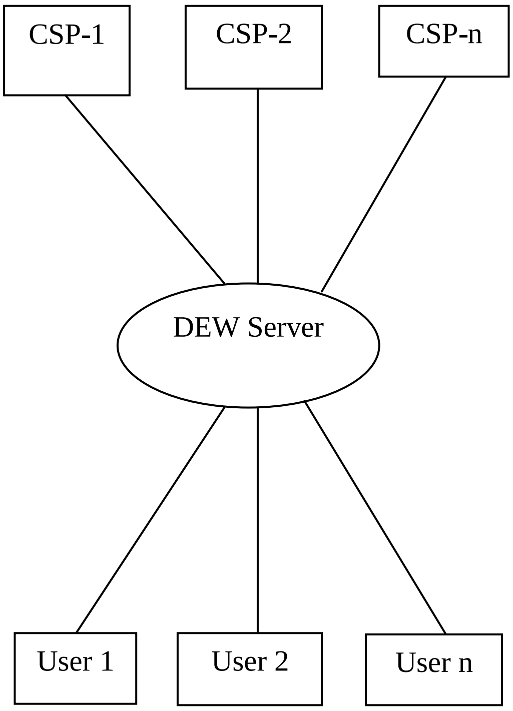

Fig.1 Block Diagram Of File Distribution On Multi- Cloud

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.

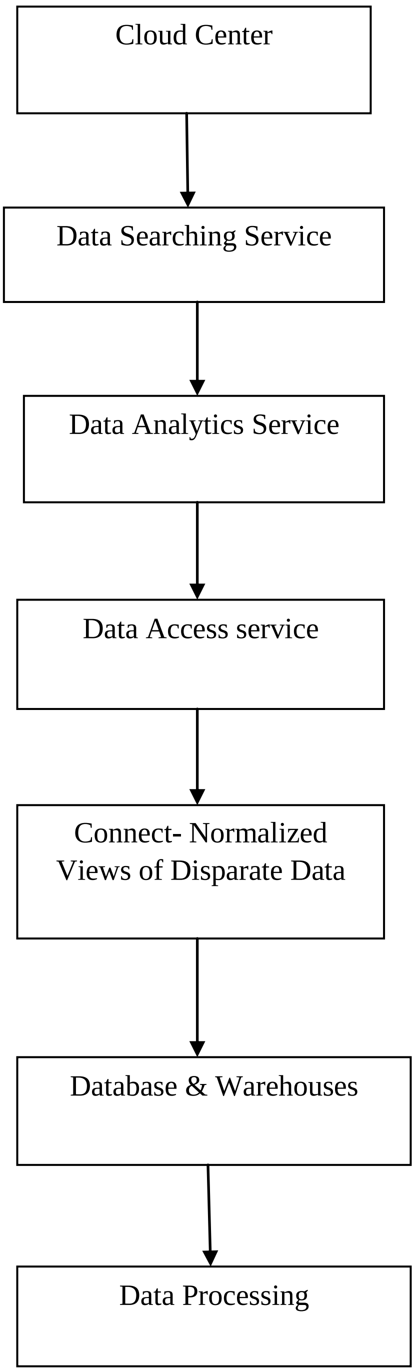

Fig.1: Block Diagram Of Cloud Storage For Big Data

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.

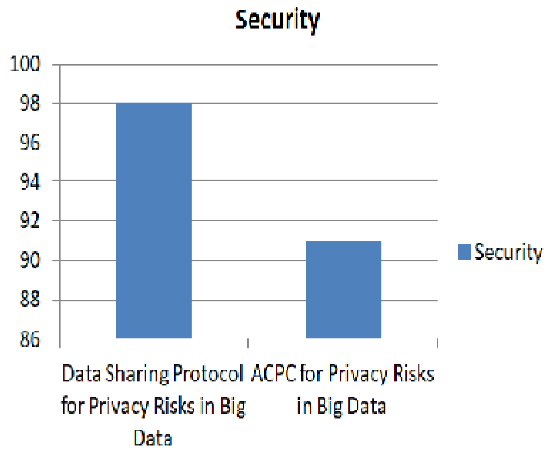

Fig.2: Security Comparison Graph

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.

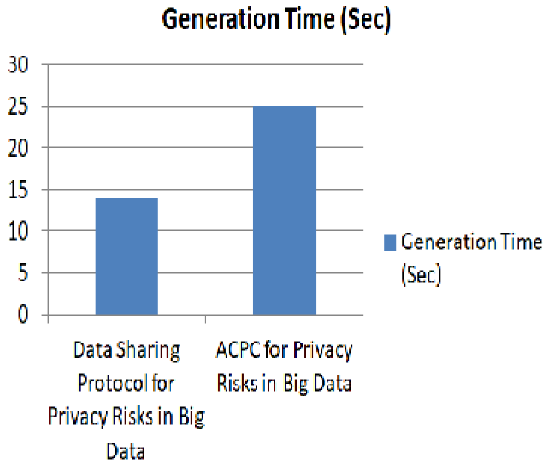

Fig.3: Generation Time Comparison Graph

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 asymmetric secure storage scheme using data sharing protocol for privacy risks in Big Data conclusion is discussed. Big data file distribution via multicloud storage. Data owner spilt his file into equal chunks and distribute them to multiple cloud storages. We analyzed security performance. Our system requires less complexity to ensure security. Also, we analyzed performance which is better than client/server. Security mechanism in our scheme guarantees the privacy of grids data in cloud storage. Encryption secures the transmission on the public channel; verified security scheme make the grids data only accessed by authorized parties. The better performance in terms of storage and computation make our scheme more practical. Hence, this analysis achieved high security and less generation time.

VI. 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.