International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-03, Issue-10, Jan 2018

Data Security Issues in Outsourced Environment: A survey

[1]Bharati P. Vasgi, [2]Dr. U. V. Kulkarni

[1]Research scholar, [2]Professor, Department of Computer Science and Engineering, SGGSIE&T, Nanded, Maharashtra, India.

[1]bharativasgi@gmail.com, [2]uvkulkarni@sggs.ac.in

Abstract - The advancement of cloud computing services is growing rapidly proportional to the rate at which organizations are outsourcing their computations and idle resources. Even if shifting to cloud platforms is an appealing trend from financial aspect, there are many other issue which has to be considered before adopting this technology. One of the important threat is privacy and security. Once the user outsource data, it is no more in his control. It is with third party cloud service provider. Aiming to give the better understanding, this paper systematically reviews the possible threats to outsourced data at various stages of data life cycle. Paper identify and classify major security threats and focuses on confidentiality as a major concern in the outsourced environment. Recent work in searchable encryption is discussed in detail.

Keywords — Data outsourcing, Security, Cloud computing, Searchable encryption, cloud storage.

I. INTRODUCTION

Cloud computing enables on demand access to shared pool of resources like hardware, software, storage, services etc. Google has introduced the concept of cloud computing. Cloud computing uses various underlined technologies like parallel computing, grid computing, distributed computing and virtualization technologies [1]. Cloud utilizes IT resources, data and application and convert into resources, which are used by customers or end users in the form of services through the Internet. The main aim of cloud computing is to use the large number of computing resources linked by the Internet and ease the job of client by relieving him from the work of maintenance. With this clients can enjoy the powerful computing ability provided by the cloud. Major IT manufactures like Amazon, Microsoft, Yahoo, Intel, and IBM [2] have started their own cloud platform.

A. Prototype of Cloud Computing

Cloud computing services are divided into three main services, Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [3] as shown in Figure 1.

- Infrastructure as a Service

Infrastructure as a service model offers large set of storage and computational environment. Through virtualization, which is a major feature provided by cloud service provider, resources like storage, computation can be divided, resized and allocated to customers as per their need [3].

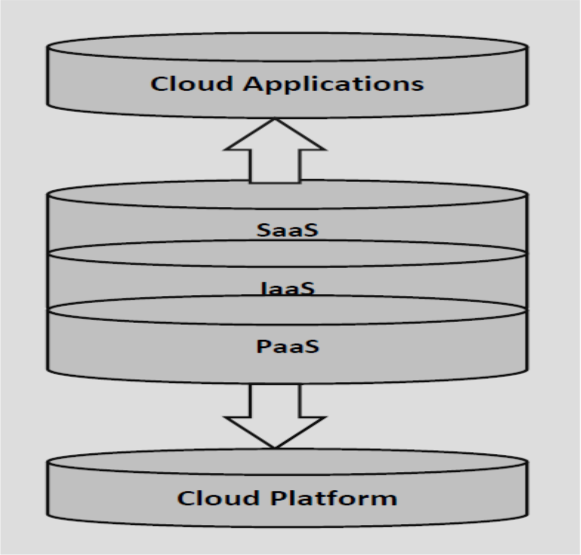

Figure 1 Types of cloud service

- Platform as a Service

Cloud systems propose vital facility by providing software as a platform on which systems run. The requirement of the hardware resources commanded by customers is made available in transparent manner [4].

DOI : 10.18231/2454-9150.2017.0083

- Software as a Service

Many of the times customers require software for some particular period of time. Buying a license copy may load on organizational budget. By using software as service users do not have to purchase the software and can use as required. Figure 2 shows top ten technology priorities. It is very clear from this that cloud computing is gaining more popularity because of its practicality [5].

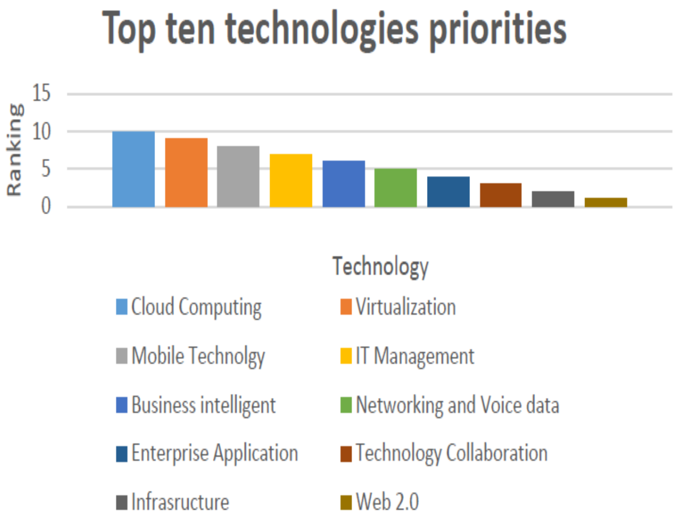

Figure 2 Top ten technology priorities

II. Significance of Security in Cloud Computing

As per the IDC (International Data Corporation) report cloud computing is the first choice among the organizations because of its flexibility and usability aspect. Figure 3 shows that the security is the major concern among all available issues in cloud computing. Hence it is very necessary to provide and discuss the problems and possible solutions [6].

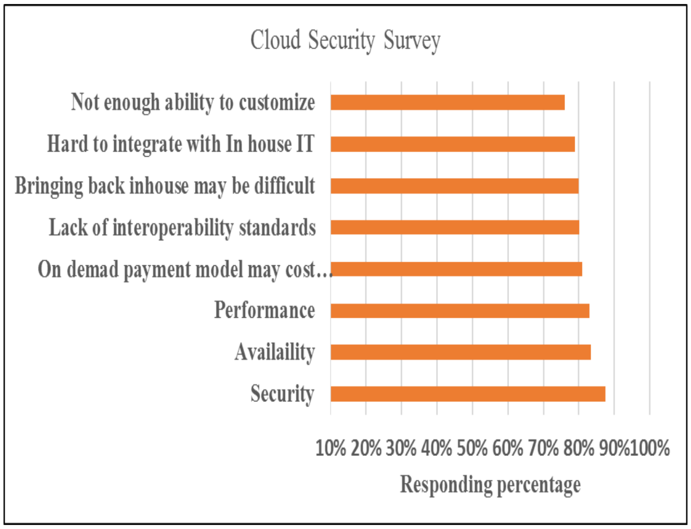

Figure 3 Cloud security survey

III. Need of Data Outsourcing

One of the commonly used outsourced data case study is EHR (Electronic health care record) Patient's health care data. Healthcare organizations (HCOs) are expected to provide new and improved patient care capabilities while healthcare cost is increasing. Several healthcare providers and insurance companies today use one or the other form of electronic medical record systems. Usually, a patient may have many healthcare service providers including primary care physicians, specialists, and therapists. In addition, a patient may enroll with various health insurance companies for different kind of insurances for e.g. dental, vision, skin [7]. Consequently, the EHR of a patient may exist at different places in the healthcare community organization. From the clinical standpoint, it is important to access the up-to-date integrated patient health information [8]. However, sharing and integration of the EHRs, that are managed by several healthcare providers is slow and costly [7] and requires effective, secure, and low cost mechanisms to share EHRs among several healthcare providers.

The requirements for storage and continuous availability of e-Health data necessitate the use of the cloud computing [9]. Cloud computing is emerging as a promising paradigm for computing. The cloud-computing model shifts the computing infrastructure to third-party service providers that is able to manage software and hardware resources with substantial cost reductions [10]. Cloud computing has shown great potential to enhance collaboration among different healthcare organizations and to fulfil the common requirements, such as scale, agility, cost effectiveness, and availability [11]. Moreover, migration of patient health records to the cloud storage relieves the healthcare providers from the infrastructure management tasks and thus reduces the maintenance cost [12], [13]. The health data can further be extracted from different databases for treatments and other analytical purposes.

Data maintained in a cloud which contains personal health records of patients like blood pressure (BP), breathing rate, blood glucose, X-rays, MRIs, Electrocardiograms (ECG), scan images, DNA reports etc. requires the proper safeguards to prevent disclosure, compromise or misuse.

IV. Threats at Various Stages of Data Life Cycle

Data is an extremely vital asset to any business. It goes through various stages and at every stage it has different threats. Figure 4 shows different phases through which data passes. Data may jump through all these stages.it is not compulsory that it passes through all phases [14].

DOI : 10.18231/2454-9150.2017.0083

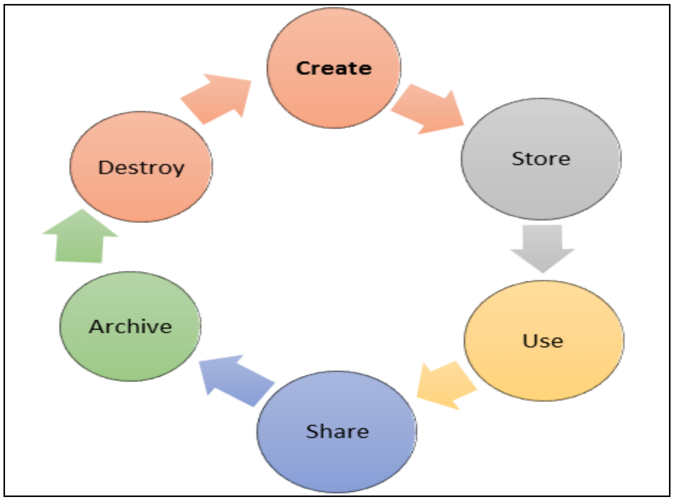

Figure 4 Data Life cycle

1. Create-When a data is newly generated or existing available data is updated.

2. Store-Data is kept or saved in some achieve. Generally this phase occurs after data passes through create phase.

3. Use-In this phase data is operated, utilized or viewed.

4. Share-Data is shared between customers or users or owners.

5. Archive-Data is back up and will not be the part of current storage system. This data is also called as historical data which may be used for any analysis.

6. Destroy-Data is permanently deleted by using physical or electronic means.

At every stage data has different threats. These threats are summarized in table 1.

Data security in cloud computing can be categorized in to following classes:

- Data –in-transit:

The data in transit mode is secured by using various encryption algorithm. Some network protocols not only provide confidentiality but also provides integrity for e.g. Hyper Text Transfer Protocol Secure [HTTPS], File Transfer Protocol Secure [FTPS] and Secure Copy Program [SCP]. These protocols are used to transfer data securely across the internet.

- Data-at-rest

To protect the data at rest encryption is the best solution. But this would prevent owner from indexing and searching of that data.

- Processing of data :

For processing the data it must not be encrypted. Data is in unencrypted form for this phase .But there are some Modern cryptographic techniques that allows to process the data without decrypting it e.g. Functional encryption, Homomorphic encryption.

- Data lineage

How data flows in the system i.e. recording the path of data is called is data lineage. Lineage is necessary for compliance purposes. It means that keeping a track of data flow in the system.

- Data Provenance

Data provenance means checking integrity of the data. It means accounting the data from its origin or creation to the current state.

- Data Remanence

Data remanence is the data present on storage device due to incomplete deletion .It can cause a serious threat to organizations. Cloud service providers are not paying much attention to this issue [15].

Data Security Life cycle stages	Data security in cloud computing
Create	Not Applicable
Store	Data at rest
Archive	Data at rest
Use	Processing of Data
Share	Data in transit, Data lineage, Data provenance
Destroy	Data Remanence

Table 1: Relation between data life cycle and classes of data security on the cloud

V. SECURITY ISSUES OF DATA IN OUTSOURCED ENVIRONMENT

Cloud Computing shifts the database and application software to the outside data centers. These data centers are geographically distributed and hence administration of software and databases may not be fully trustworthy [15]. Trust management: Trust management is defined as reliance on the integrity, strength, ability and surety of a person or thing. Entrust your data on to a third party who is providing cloud services is an issue. Some of the important security issues are shown in figure 5.

i. Data location and Relocation: Cloud technology proposes a high data mobility because of geographically separated data centers. Customers are not aware of location of their data [15].

DOI : 10.18231/2454-9150.2017.0083

ii. Data integrity: Data integrity is an important feature which must be implemented by any secure system. This ensures that customer’s valuable data is not tampered during the course of agreement [15].

iii. Data recovery: There can be the incidences like server failure or natural calamity at data center. This may cause loss of customer’s important data. To avoid these problems customer’s data is maintained by implementing various mirror servers. So even though any server is damaged data can be recovered from its mirror location. Cloud users can also keep a backup of critical data on a local computer.

iv. Data Backup: Many cloud services now a days allows their users to take back up of their data through regular downloads. Many service providers are also allowing data dumps on storage media.

v. Data Portability and Conversion:
Many times customers have issues like changing service provider which may cause difficulty in shifting the data. Converting and shifting data is largely dependent on service provider’s retrieval format [16].

vi. Access Control
Access control consist of both authentication and authorization. Cloud service provider implies very poor authentication mechanism like using username and password. Advanced authentication techniques must be implemented by cloud service provider in order to sustain today’s attacks.

vii. Confidentiality
Resources must be accessed by authorized entities. This important feature of secure system is called as confidentiality. This property of secure system makes sure that resources are accessed for reading, writing or viewing by authorized people only Data stored on the cloud could be compromised by the cloud computing provider, or by other customers of the provider, who could be competitors. Customers are unaware of the data processing practices that take place in the cloud.

Cloud providers employ data mining techniques, which extract information or patterns from large amounts of data, to improve the user experience and provide users with better service. This raises ethical and privacy concerns, especially since users are storing private and sensitive data on the cloud such as private photos.

viii. Data deletion: It is another issue of concern. Since data on cloud is scattered geographically it is very difficult to guarantee that all copies of data are removed. Hence it is difficult to assure deletion of data [16].

ix. Data Availability
Availability of data is an important building block of any secure system. The strength of cloud computing system will be at sake if it is not highly available. Many customers will not adopt cloud infrastructure if it is not highly available [16].

VI. CONFIDENTIALITY SOLUTION

The important techniques used to maintain data confidentiality is shown in figure 6.

A. Oblivious RAM (ORAM)

Oblivious RAM (ORAM) technique proposed by Goldreich and Ostrovsky [17], permits a customer to hide its access pattern to far away stored data by periodically shuffling and enciphering data whenever they are retrieved. This is very important because the order of storage locations read by the client, tampers important sensitive information through statistical inference [17].

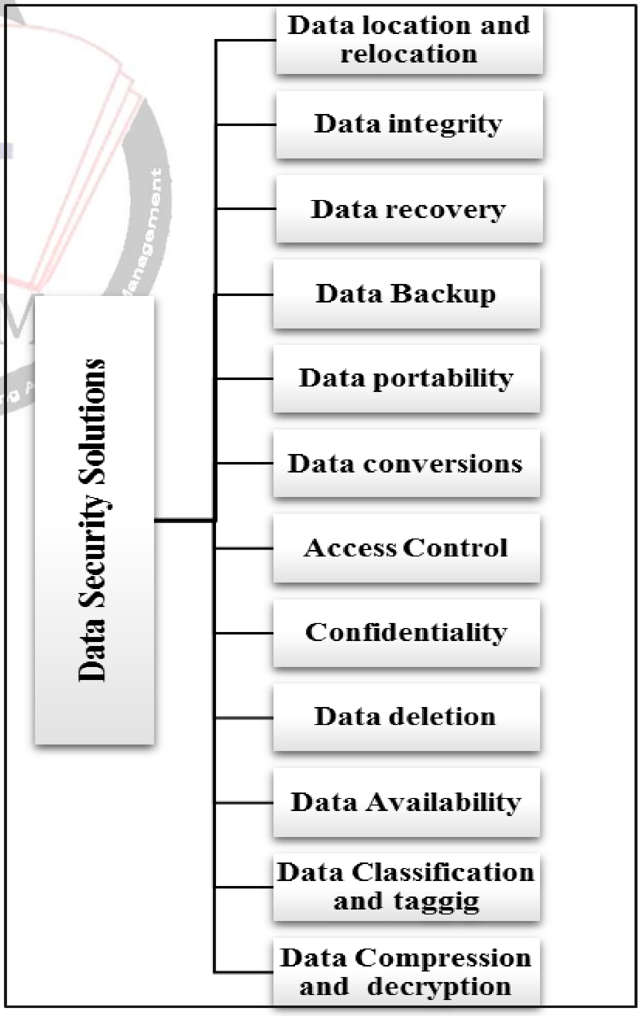

Figure 5 Data Security solutions in outsourced environment

DOI : 10.18231/2454-9150.2017.0083

B. Homomorphic Encryption

Homomorphic encryption systems plays very important role in modern cryptography. By using this techniques we can operate on encrypted data without deciphering it. Whenever we decrypt the result of any operation, it is observed that it is same as if we have carried out the operation on raw data.

Definition: An encryption is said to be a homomorphic, if it is possible to compute Enc (f (x, y)), if Enc(x) and Enc(y) is given. Here f indicates additive or multiplicative operation without use of private key. Depending upon the operations that allows access to raw data Homomorphic encryption is classified as additive Homomorphic encryption [18], [19] and the multiplicative Homomorphic encryption [20] , [21] .

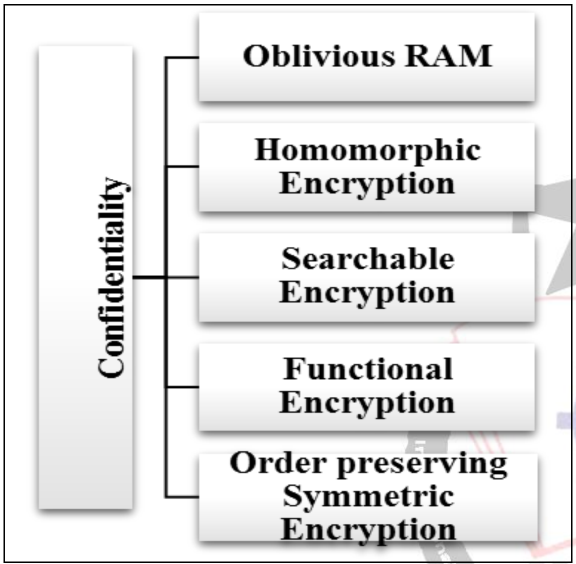

Figure 6 Confidentiality approaches

C. Searchable Encryption (SE)

In SE schemes, a user can outsource a collection of encrypted data to the server while maintaining the ability to search them. From the aspect of security, the privacy of documents and keywords is maintained. The searchable encryption has two variations Searchable Symmetric Encryption (SSE) and public key encryption with keyword search (PEKS). SSE is based on the private key primitive. It allows only the private key holder to produce cipher texts and to create trapdoors for search. PEKS, on the other hand, is related to the public key primitive. It enables a number of users who know the public key to produce cipher texts .The users having private key can only create trapdoors required for search [22]. Fast retrieval using hash based map structure is suggested in [23].

SSE and PKSE has schemes like
1. Single keyword search
2. Fuzzy keyword search
3. Conjunctive keyword search
4. Ranked and verifiable keyword search

Table 2 summarizes shows some of the recent advancements in the area of searchable encryption. All the existing techniques works on textual data. Future work in the area of searchable encryption can be handling data in various formats like image, video etc.

D. Functional Encryption

In this scheme with the help of decryption key user can learn the function of the encrypted information. In other words in this technique the authorized user holding master key can generate a key Sk which helps in the computation of the function F(k) on encrypted data.

More accurately, the decryptor can calculate F (k; x) by using key Sk provided he knows the encryption of x. This indicates that the security of the system ensures that no one can learn additional information about x [24].

E. Order preserving Symmetric Encryption (OPSE)

Order preserving encryption (OPE) is a technique to make efficient comparisons on encrypted data without decrypting it. It is a deterministic symmetric encryption which produces cipher text which maintains numerical order of the plaintext. An order-preserving symmetric encryption (or OPE) technique is a deterministic symmetric encryption algorithm. The ciphertext produced by this technique ensures that numerical ordering of plaintext is maintained [25]. Initially OPE was suggested from database community Agrawal et al. [26] in 2004 which supports range queries efficiently over encrypted data.

OPE is having various schemes like random order preserving function (ROPF), property preserving order preserving function (POPF), modular order preserving encryption (MOPE) [25].

VII. CONCLUSION

With the invention of cloud computing application software, databases are shifted to large data centers. These data centers may be located geographically at large distance. Hence management of data and services may not be fully trustworthy. In this paper we have studied need of data outsourcing and importance of storage security in cloud computing environment. Various stages of data life cycle and threats at these various phases are discussed. Some of the important security solutions related to data outsourcing are presented. In this paper we discuss modern cryptographic techniques that are used to secure data e.g. ORAM, Homomorphic encryption etc. The recent work on searchable symmetric encryption is discussed in detail. The future work will be to enhance searchable encryption to achieve retrieval efficiency in terms of time and space.

DOI : 10.18231/2454-9150.2017.0083

Sr.No	Methodology used	Advantages	Limitations	year
1	Inverted index Message digest [27]	1.Support score dynamics 2.Use of OPSE 3.Supports ranked search 4.Maintains keyword and data privacy 5.Authentication of ranked search 6.Efficient	1. Only single keyword search. 2. Does not handle addition and deletion of files. 3. Little relevance score info leakage. 4. Server learns the order of relevance scores. 5.Two way trip for retrieval	2012
2	1.Cloud service provider supports in partial decipherment [28]	1.Semantically secure against chosen plaintext attack	No support for range queries	2012
3	1.Uses vector model, 2.Secure KNN approach [29]	1.Multikeyword ranked search 2.Privacy-preserving 3.Efficiency 4.Supports score dynamics	Integrity is not verified in untrusted environment	2014
4	1.Use of vector space model 2.cosine similarity 3. Secure Index tree [30]	1.Multikeyword 2.Ranking 3.Result verification 4.Balance between search precision and privacy	Multiuser environment not supported	2014
5	1.Use of semantic tree 2.Term similarity tree [31]	1.Proposes smart keyword-based semantic search scheme 2.supportis verification of completeness	1.Does not support semantic search 2. conjunctive keywords	2014
6	1.Introduced grouping-based construction [32]	1.Address search pattern leakage issue 2. stronger security guarantee	1.Does not handle relevance ranking 2.Range queries	2014
7	1.use of balanced binary tree based index structure 2..Depth first search technique 3.Use of parallel search [33]	1.Supports dynamic update operation 2.Supports Multikeyword	1. Data owner responsible for updating index, rather it should be the part of CSP 2.Index tree is plaintext 3. Revocation of key is biggest challenge	2015
8	1.Clustering method is used 2. Uses semantic relationship between plain documents [34]	1.Search time is linear instead of exponential 2.Check Integrity and privacy requirement 3.Supports multikeyword	Queries are keyword based	2015
9	1 Use of attribute based encryption 2.Adopted secure KNN approach [35]	1. Confidentiality of Documents and Index 2. Trapdoor Privacy 3.Trapdoor Unlink ability 4.Concealing Access Pattern of the Search User	No support for range queries	2015
10	1Use of attribute based encryption 2.Proxy reencryption 3.Lazy reencryption [36]	1.Handles user revocation 2.Supports Fine grained search 3.Secure against chosen keyword attack 4.Supports authenticity check 5.Handles conjunctive keyword search	1. No support for multikeyword search 2.No ranking	2016
11	1.Uses sub directory classification 2.Uses secure KNN approach [37]	1. Supports multikeyword 2. Handles relevancy 3.Supports fine grained access control	1.Cannot work in multiuser environment 2.Serachable encryption is not highly scalable	2016

Table 2: Comparative study of advancements in Searchable Encryption

REFERENCES

[1] K. Meng. A Walk in the Cloud:Uncovering Cloud Computing. pp.12-14, 2008.6.16

[2] G. Boss, P. Malladi, D. Quan, L. Legregni, H. Hall. Cloud Computing:IBM White Paper.http://download.boulder.ibm.com/ibmdl/pub/software/dwes/hipods/Cloud_computing_wp_final_8Oct.pdf, 2007.

[3] J. Geelan. Twenty One Experts Define Cloud Computing.Virtualization. Electronic Magazine, http://virtualization.syscon.com/node/612375, 2008.8.

[4] C. L. Li, Z. H. Deng. On the Value of Cloud Computing. No.4, pp.42-46, 2009

[5] Shuai Z; Shufen Z; Xuebin C; Xiuzhen H; (2010), "Cloud Computing Research and Development Trend",

2nd International conference on Future Networks, 2010. ICFN ' 10. pp 23, 22-24 Jan 2010.

[6] Chang, L, Ti ; Chin L; Chang, A.Y.; Chun J, C;(2010), " Information security issue of enterprises adopting the application of cloud computing", IEEE 2010 Sixth International Conference on Networked Computing and Advanced Information Management (NCM),pp 645, 16-18 Aug. 2010.

[7] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in Proc. 3rd IEEE Int. Conf. Cloud Comput., Miami, FL, USA, Jul. 2010, pp. 268–275.

[8] R.Wu, G.-J. Ahn, and H. Hu, "Secure sharing of electronic health records in clouds," in Proc. 8th IEEE Int. Conf. Collaborative Comput., Netw., Appl. Work-sharing, 2012, pp. 711–718.

[9] S. P. Ahuja, S. Mani, and J. Zambrano1, "A survey of the state of cloud computing in healthcare," Netw. Commun. Technol., vol. 1, no. 2, pp. 12–19, Sep. 2012.

[10] B. Grobauer, T. Walloschek, and E. Stocker, "Understanding cloud computing vulnerabilities," IEEE Security Privacy, vol. 9, no. 2, pp. 50–57, Mar. 2011.

[11] R.Buyya,J.Broberg, and A.M.Goscinski ,Cloud computing:Principles and paradigms,vol 87.John Wiley & Sons,2010.

[12] S. Yu, C.Wang, K. Ren, and W. Lou, "Achieving secure, scalable and finegrained data access control in cloud computing," in Proc. IEEE Infocom, Mar. 2010, pp. 1–9.

[13] B. Horowitz. (2012, Dec. 20). "Cloud comput. brings challenges for health care data storage, privacy," [Online]. Available:http://www.eweek.com/c/a/Health-Care-IT

[14] Data Security Life Cycle. Available : https://securosis.com/blog/datasecuritylifecycle-2.0

[15] Sathyanarayana, T. V., and L. Mary Immaculate Sheela. "Data security in cloud computing." In Green Computing, Communication and Conservation of Energy (ICGCE), 2013 International Conference on, pp. 822-827. IEEE, 2013.

[16] 16. Dinadayalan, P., S. Jegadeeswari, and D. Gnanambigai. "Data security issues in cloud environment and solutions." In Computing and Communication Technologies (WCCCT), 2014 World Congress on, pp. 88-91. IEEE, 2014.

[17] O. Goldreich and R. Ostrovsky. Software protection and simulation on oblivious rams. J. ACM,1996.

[18] Pascal Paillier. Public-key cryptosystems based on composite degree residuosity classes. In 18th Annual Eurocrypt Conference (EUROCRYPT'99), Prague, Czech Republic , volume 1592, 1999

[19] Julien Bringe and al. An Application of the Goldwasser-Micali Cryptosystem to Biometric Authentication, Springer-Verlag, 2007.

[20] R. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public key cryptosystems. Communications of the ACM, 21(2) :120-126, 1978. Computer Science, pages 223-238. Springer, 1999

[21] Taher ElGamal. A public key cryptosystem and a signature scheme based on discrete logarithms. IEEE Transactions on Information Theory, 469-472, 1985.

[22] Wang, Yunling, Jianfeng Wang, and Xiaofeng Chen. "Secure searchable encryption: a survey." Journal of communications and information networks 1, no. 4 (2016): 52-65.

[23] B. P. Vasgi and U. V. Kulkarni, "A secure and effective retrieval using hash based mapping structure over encrypted cloud data," International Journal of Electrical Electronics and Computer Science Engineering, vol. 4, no. 4, pp. 65-74, 2017.

[24] Boneh, Dan, Amit Sahai, and Brent Waters. "Functional encryption: Definitions and challenges." Theory of Cryptography (2011): 253-273.

[25] Boldyreva, Alexandra, Nathan Chenette, and Adam O'Neill. "Order-Preserving Encryption Revisited: Improved Security Analysis and Alternative Solutions." In CRYPTO, vol. 6841, pp. 578-595. 2011.

[26] Agrawal, Rakesh, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. "Order preserving encryption for numeric data." In Proceedings of the 2004 ACM SIGMOD international conference on Management of data, pp. 563-574. ACM, 2004.

[27] Wang, Cong, Ning Cao, Kui Ren, and Wenjing Lou. "Enabling secure and efficient ranked keyword search over outsourced cloud data." IEEE Transactions on parallel and distributed systems 23, no. 8 (2012): 1467-1479.

[28] Liu, Qin, Guojun Wang, and Jie Wu. "Secure and privacy preserving keyword searching for cloud storage services." Journal of network and computer applications 35, no. 3 (2012): 927-933.

[29] Cao, Ning, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. "Privacy-preserving multi-keyword ranked search over encrypted cloud data." IEEE Transactions on parallel and distributed systems 25, no. 1 (2014): 222-233.

[30] Sun, Wenhai, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, and Hui Li. "Verifiable privacy-preserving multi-keyword text search in the cloud

DOI : 10.18231/2454-9150.2017.0083

supporting similarity-based ranking." IEEE Transactions on Parallel and Distributed Systems 25, no. 11 (2014): 3025-3035.

[31] Fu, Zhangjie, Jiangang Shu, Xingming Sun, and Nigel Linge. "Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data." IEEE Transactions on Consumer Electronics 60, no. 4 (2014): 762-770.

[32] Liu, Chang, Liehuang Zhu, Mingzhong Wang, and Yu-an Tan. "Search pattern leakage in searchable encryption: Attacks and new construction." Information Sciences 265 (2014): 176-188

[33] Xia, Zhihua, Xinhui Wang, Xingming Sun, and Qian Wang. "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data." IEEE Transactions on Parallel and Distributed Systems 27, no. 2 (2016): 340-352.

[34] Chen, Chi, Xiaojie Zhu, Peisong Shen, Jiankun Hu, Song Guo, Zahir Tari, and Albert Y. Zomaya. "An efficient privacy-preserving ranked keyword search method." IEEE Transactions on Parallel and Distributed Systems 27, no. 4 (2016): 951-963.

[35] Li, Hongwei, Dongxiao Liu, Yuanshun Dai, Tom H. Luan, and Xuemin Sherman Shen. "Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage." IEEE Transactions on Emerging Topics in Computing 3, no. 1 (2015): 127-138.

[36] Sun, Wenhai, Shucheng Yu, Wenjing Lou, Y. Thomas Hou, and Hui Li. "Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud." IEEE Transactions on Parallel and Distributed Systems 27, no. 4 (2016): 1187-1198.

[37] Li, Hongwei, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, and Xuemin Sherman Shen. "Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data." IEEE Transactions on Dependable and Secure Computing 13, no. 3 (2016): 312-325.

DOI : 10.18231/2454-9150.2017.0083