IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

Data Security in Cloud Computing: A Survey

Shaziya Banu[1*], Gopal K Shyam[2]

[1,2]School of Computing and Information Technology, Reva University, Bangalore, India

Corresponding Author: shaziyabanu8888@gmail.com, mobile no.: +91 – 9742202321

DOI: https://doi.org/10.26438/ijcse/v7si14.245251 | Available online at: www.ijcseonline.org

***Abstract*—** Cloud computing is increasing its importance due to the services and IT resources provided by the cloud service providers. There are numerous advantages in cloud computing like highly scalable, low cost and high availability. But the data security and privacy is a major concern in cloud computing. Sharing the resources and storing the data in cloud is the major application in cloud computing. To protect the data in cloud, against unauthorized access, modification and denial of service etc. The biggest challenge in cloud is storing and sharing the sensitive data. In this paper, we have discussed symmetric algorithms, asymmetric algorithms, hash algorithms for security purpose and comparison of algorithms for securing data in different deployment models of cloud. In symmetric we have discussed AES, Fully Homomorphic Encryption, Blowfish, GLEnc, and asymmetric algorithms are DIFFE-HELLMAN, QKD-NAE Technique, RSA Cryptosystem, Identity Based Encryption. In hash few of the algorithms that has been discussed in this paper are MD5, SHA.

***Keywords*—**Cloud Computing, Data Security, Symmetric Encryption, Asymmetric Encryption, Hash function.

I. INTRODUCTION

Cloud computing is a technology that delivers IT resources as a service (IAAS, PAAS, SAAS) to external customers through Internet. Cloud delivery models provided by cloud service providers are prepacked combination of IT resources.

A. Cloud Deployment Models

A cloud deployment model can be distinguished by its ownership, size and access. Based on these representations there are four cloud deployment models: public cloud, community cloud, private cloud, hybrid cloud.

1) public cloud: public cloud is accessible publically which is owned by the third party cloud provider by internet browser via web to cloud consumers. Cost is commercialized via other approach like advertisement. The on-going maintenance and creation of the public cloud IT resources is at the helm of cloud providers. Data is not secured in public cloud. Public cloud has a biggest security threat among the cloud deployment models.

2) Community cloud: Community model is as similar to public model except its access is collectively sharing infrastructure between multiple organization. Community cloud ownership may be pooled by community members or by a third party cloud providers. Members of community cloud share responsibility among them.

3) Private cloud: Private cloud is handled by a single organization itself. Private cloud enables an organization to use IT resources by means of centralizing access either from different parts and locations or departments of an organization. The greatest advantage in private is that its easier to manage security, maintenance and upgrade. The resources and applications are square measure managed by the organization itself.

4) Hybrid cloud: A hybrid cloud is a combination of different (two or more) cloud deployment models (private, public, community). These models remains unique but are bounded together by some standards. One such hybrid model is when the consumer want to store the secured data then he can store it in private cloud and less secured cloud services to a public cloud.

Cloud providers offers services through internet at low cost and high availability. All these offerings helps in achieving scalability to meet needs of cloud consumers at any given time. The pay-per-use model characteristic of cloud enables the cloud consumers to utilize the resources as per their requirement and usage. With all these benefits, cloud has some disadvantage; one of them is data security and privacy. To overcome these disadvantages of data security and privacy we can utilize the best algorithm. These algorithms should provide security against unauthorized access, modification and denial of service etc from unauthorized persons and attackers.

Cryptography helps in achieving secure communication of data [16].There are 5 primary functions of cryptography (Fig.1).

1) Confidentiality- No one reads the message except the authorized person.
2) Authentication- The process or action of verifying the identity of a user.
3) Integrity- Assuring the receiver, that the received message is same as original message.
4) Non-repudiation- A mechanism to ensure that the message has been sent by the sender.
5) Key exchange- key exchange is the method in which both the sender and receiver shares the crypto keys.

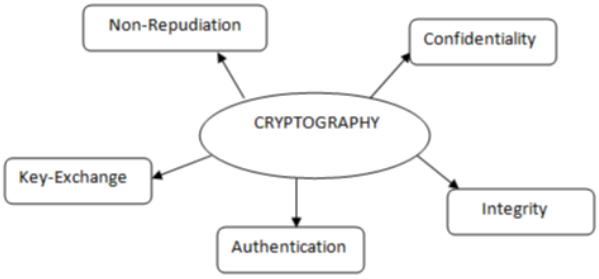

Fig. 1. Cryptography Functions

In order to achieve these 5 functionality of cryptography, there are several types of cryptography algorithms for security. There are 3 types of cryptography classified based on the key.

1) *symmetric encryption:* In symmetric encryption both encryption and decryption uses a single key, hence also called secret key algorithm (see fig 2), it offers privacy and confidentiality. e.g, AES, BLOWFISH, Fully Homomorphic Encryption.

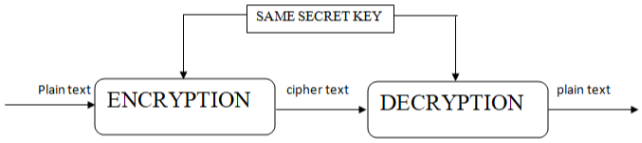

Fig. 2. Symmetric Encryption

2) Asymmetric Encryption: In asymmetric encryption two keys are involved, one key for encryption, another key decryption it is also called public key cryptography (see fig 3), capable of providing authentication, non-repudiation and key exchange. e.g, RSA, DIFFE-HELLMAN, ECC.

3) Hash function: Hash algorithm provides integrity to verify input messages through digital signatures with key exchange, message authentication. Hash function provides a digital fingerprint (see fig 4) uses mathematical transformation to ”encrypt” information. Mostly used when message integrity has to be maintained. e.g: MD, SHA, RIPEMD, HAVAL.

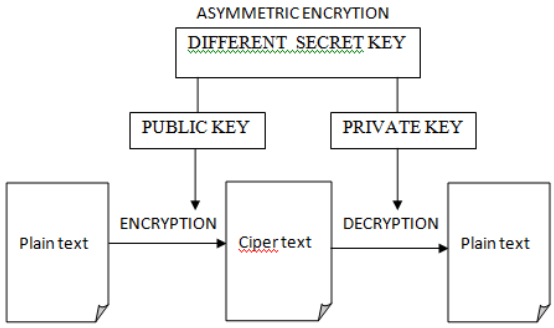

Fig. 3. Asymmetric Encryption

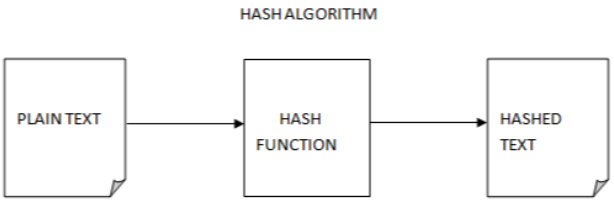

Fig. 4. Hash Function

The rest of the paper is organized as follows, Section I contains the introduction of cloud computing technology, Section II contain the related work of risk and challenges faced in data security. Proceeding with section III issues in data security. Followed by section IV with review of literature survey, section V is the concludes the paper with future work.

II. RELATED WORK

A. *Risks and Challenges in Data Security*

Risks are caused due to the threat agents. Threat agent is an entity that poses a threat, threats can originate either externally or internally, caused by human beings or software programs. Data confidentiality and integrity can be lost in cloud computing due to guessing password attack, account hijacking attack [1].

Several risks are associated with the data in cloud computing. This section is related to the discussion of data security associated with virtualization, multi-tenancy and storage in cloud.

1. Virtualization: Virtualization is a basic premise in cloud computing which delivers the core values of cloud computing. Virtualization is a technique, in which virtual instances of IT resources are created. With the physical IT resources, the virtualization layer creates multiple copies of virtual images of themselves in order to utilize the underlying processing capacities among many users. Risk

associated in virtualization is during allocation and de-allocation of IT resources. During allocation, data might be at risk, when he data is not erased in the memory before reallocation of the resource to the next user. This can be resolved by properly authenticating data before de-allocation of the resources.

2. Multi-tenancy: Multiple users sharing same memory, storage and IT resources. The multi-tenancy technology which is based on the virtualization technology helps the cloud service providers to provide service for multiple cloud service consumers. Resource pooling allows cloud providers to pool large-scale IT resources to provide service to several cloud consumers, which is commonly achieved through multi-tenancy. In such cases, there is always a risk of leaking the data to other tenants. Hackers can also misuse the data of the tenants this can be avoided by properly authenticating the tenants before accessing the data.

3. Storage: Data storage in a public cloud is again a risk in cloud, which becomes a target for the hackers.
It is very difficulty to secure data and ensure the safety of data when the computers are linked in series and when clients are associated with me. When it comes to the matter of securing data in cloud , it becomes a biggest challenge for the cloud service providers to secure the data. The major challenges involved for cloud service providers are

1. Lack of appropriate governance: Cloud service providers have full control over the data. When the control is under the cloud service providers, then there is a threat for the authorized parties being compromised its security with the cloud service providers, which may lead to unauthorized data access and application of the resources. Service Level Agreements comes as a concern for this, but another threat occurs when this SLA are not in place with the cloud service providers.

2. Insecure or incomplete data: When the client requests for the cloud service providers to delete the data partially or completely from the cloud, the problem arise here is whether the data can be deleted partially from the cloud in the form of data segments with accuracy.

3. Data interception: In cloud computing data is segmented and distributed in transit. This causes more threats due to the exposure of attacks and makes unreliable for the computing technology and, in particular, sniffing and spoofing, third party attacks.

B. Issues in Data Security
There are two states of data in cloud which causes threat in cloud security [2].
1. Data at Rest: The data is said to be at rest when he data is stored in cloud that can be accessed through internet. The data stored in the cloud can be either the backup data or the real time data. It is very difficult for the organizations to secure data which is at rest.

2. Data in Transit: The data which is moved in and out of the cloud is called as the data in transit. The data transferred to the cloud can be database stored on the cloud or any data. The data transmitted in and out can be very sensitive data like user names, passwords, confidential documents and can be encrypted sometimes. Data in transit is more riskiest than the data at rest because when data travels from one location to another may have more chance of attacks than the data at rest. In order to secure data, best strategy is encryption.
Securing data at rest and data at transit involves different strategies. Confidentiality and integrity of the data can be achieved by different protection mechanisms, procedures and processes.

III. LITERATURE SURVEY

Different algorithms have been used for securing data in cloud. Based on the parameters like Confidentiality, Authenticity, Integrity, Authorization, Non-repudiation, Freshness algorithms can be differentiated and used for data security based in the required parameters.
To provide security to cloud many algorithms are designed, based on Symmetric, Asymmetric, Hash function.

1) Symmetric Encryption
A) AES: Advanced encryption standard is the most efficient symmetric algorithm used for security purpose since it offers more speed than the other algorithms. Considering HEROKU as an example for paas (Platform as a service) the three authors proposed this model for the security purpose. Bih-Hwang Lee, Ervin Kusuma Dewi and Muhammad Farid Wajdi [3] has proposed AES 128 under HEROKU cloud to encrypt the data. It has 4 steps.

- Substitute Bytes: In this step, replace each byte of the input data with another byte in the substitution table(s-box) and in his s-lookup table which is 8 bit fixed size is replaced by the entry of each byte state.
- Shift Rows: After replacing, the bytes in each row of the s-box, then these rows are shifted cyclically to the left and for each row the number of bytes shifted will differ. Do not confuse “imply” and “infer”.
- Mixing Columns: Each column of the state are multiplies of fixed polynomial.
- The AddRoundKey: After mixing columns, each byte of the state is then combined with a byte of round subkey using XOR operation($\oplus$) .

HEROKU is an integrated tool, which supports some of the programming languages like Java, Python, Ruby. It runs on Dynos app. HEROKU’s infrastructure provides some of the best features for the developers like firewall, spoofing, and

sniffing protection, port scanning, and also DDOS mitigation. It helps in securing the data in cloud , the data stored can be any type like database. It also helps in encryption of the stored data through customer's app for security purpose.

B) Fully Homomorphic Encryption: Ahmed and Mohammed Dafir [4] proposed a fully homomorphic encryption scheme does not follow the Euler's theorem instead it calculates the modulus of two big prime factors. It is a quadruplet of algorithms(Gen, Enc, Dec, Eval).

- *Gen(□):* Key generation algorithm is utilized for key generation which takes a security parameter (□) as a input and produces a public and secret keys(*pk*,*sk*) as output .
- *Enc(m,pk):* Its an encryption algorithm, which plaintext and a public key as an input and gives output as a ciphertext *c*.
- *Dec(c,sk):* The output of encryption algorithm i.e the cipertext as input along with a secret key and outputs a plaintext m.
- *Eval (C,c1,…,cn):* It takes as input a circuit C and ciphertexts c1,..,cn and verifies as in

$$Dec(Eval(C,c1,\ldots,cn),sk)=c(m1,\ldots,mn) \quad (1)$$

Anyone can evaluate Eval, since it does not require the secret key sk. Fully homomorphic encryption is probabilistic, noise free which consists of four main algorithms:

- Key Generation: generate two prime numbers p and q, then calculate n=p.q, where p and q should be kept secret.
- •Encryption: let $m \in \mathbb{Z}/p\mathbb{Z}$ be a clear text. To encrypt *m* first generate randomly an integer r, then cipertext of *m* is

$$C= m^{r(p-1)+1} \ mod \ n. \quad (2)$$

- Decryption: let $C \in \mathbb{Z}/n\mathbb{Z}$ be a cipertext, then recover the plaintext m by calculating m=C *mod* p.
- Evaluation: let m_1 and m_2 be two clear texts and $C_1=Enc(m_1)$ and $C_2=Enc(m_2)$ be cipertext respectively, then

$$C_{add}= C_1+C_2 mod \ n = Enc(m_1) + Enc(m_2) mod \ n = Enc(m_1 + m_2), \quad (3)$$

$$C_{mult}= C_1.C_2 mod \ n = Enc(m_1) \ . \ Enc(m_2) mod \ n = Enc(m_1 \ x \ m_2). \quad (4)$$

C) Blowfish Algorithm: Blowfish was designed by Bruce Schneier in 1993, can be used instead of existing encryption algorithms[12]. Blowfish comes under symmetric block cipher which can be used instead of DES. It takes a variable-length key, starting from 32 bits to 448 bits, which is considerably better to use for both domestic and commodity. After the verification of the algorithm its popularity increased gradually and hence became a strong encryption algorithm. Blowfish is license-free, hence it is available free for all users.

D) GLEnc: The GLEnc encryption algorithm encrypts the data before entering data into the cloud storage. Since it provides encryption at the starting state during entry of data into the cloud, its the highest security algorithm for public cloud storage.

In GLEnc encryption algorithm[11] the plain text values are first counted and converted to ASCII decimal values and then these ASCII values are converted to their binary format. In this process of encryption, the values undergo splitting and dividing with bit conversion. In GLEnc encryption algorithm three keys are generated at different stages. XOR is performed for the generation of third key. Then it is changed to decimal format and to provide the cipher textual content, it is converted into ASCII code.

A security analysis tool called ABC Hackman is used for GLEnc security level analysis. This tool performs security analyses on different security levels of proposed and existing techniques. This security tool is installed in the cloud server for the analysis of security levels in different proposed and existing techniques such as DES, 3DES and Blowfish. In the cloud server when the hackman attacks the encrypted text, it can attack in different forms like dictionary and brute force attack to retrieve the original text. At the end of retrieval, to find out the percentage of original text retrieved, the hackman compares the plain text with the retrieved text. The security level of the proposed algorithm is measured by comparing the percentage of the plain text with the retrieved data.

2) Asymmetric Encryption

A) Modified Elliptic Curve Cryptography(MECC): Thangapandiyan, rubesh and sakthidasan[5] proposed Modified Elliptic Curve Cryptography provides security for sensitive data of the bank, then the data is saved in the bank server. In this admin is the data owner who divides data into sensitive and non-sensitive parts. The sensitive data is encrypted. All he partakers of cloud computing will have a key.

To encode the data, attributes of the data along with marketing admin key, insurance admin key, loan admin keys and customers account number has been utilized.

When the admin transfers the encrypted sensitive data and non-sensitivity data to cloud then the requestors gets customer details from the cloud based on their requirement for further process. This requested data has been transferred by cloud service providers to the admin which is hinted to the respected customers. With the Approval of customers, the admin can access the data in the form of an encrypted key otherwise the admin can't access the data from the cloud.
The decrypt part is handled by the admin and when matches the requirement of the requestor then only the decision of

admin is been transferred to the cloud and then the requestors get the decision from the cloud.

B) QKD-NAE Technique: Quantum key distribution based Non-Abelian Encryption provides both transmission and data security. Thangapandiyan, Rubesh, Sakthidas[6] proposed methodology is for the application of personal health record where the transmission level security is maintained by utilizing Diffe-Hellman(DH) based scheme and mechanism i.e. secret exchange mechanism and key distribution scheme to generate keys which is based on the index of qubits. Non-abelian encryption and decryption for securing data during storage and while accessing the data from the cloud. The proposed system consists of 4 modules, quantum key distribution, secret generation based on access policy verification, data encryption and data decryption. first the user has to register for login access so that only the authorized users can access the data. The records are maintained by the admin to provide restricted access to the valid user.

- QKD: String of secret bits known as Qubits, these secret bits are generated by the admin from the quantum key for encryption and decryption. Quantum key is used by both the user and admin for data storage and access data from cloud. The data can be accessed only through a onetime password which is generated for both user and admin. When the user requests to access the data from cloud, then the admin sends qubits to the user. Using these qubits, the user generates a set of strings and then again forwards these strings to the admin. The qubits are then matched based on its index by admin, if it matches, then only key is used at both ends for encryption and decryption by denoting it either 0 or1.
- Access Policy Verification: After the key is generated and distributed, then the admin generates a distinct key by utilizing Diffie-Hellman algorithm(it is present at cloud for secured authentication purpose, this technique does not contain any key for data storage and exchange) for security purpose to create access policies for each cloud user. Based on the user destination, the secret key is generated for each and every cloud user.
- Data Encryption: The original data is split into multiple fragments and encrypted it by using NAE technique. Here the quantum key is produced by fulfilling the non-abelian property and once the public key is generated it is then forwarded to the user for decryption of the data. Based on the XORed values encrypted ciphers are generated to create encrypted files, then transfer these files to different servers in the cloud. Finally after signature generation, with the NA decryption technique, the fragmented data is decrypted separately. when the signature matches with the user's signature (verification of signatures is performed by hashing function), the data can be decrypted.
- Data Decryption: After receiving the data from multiple cloud servers, suitable signatures are generated for the corresponding data. The quantum key is the master key and the random key is generated based on the master key through non-abelian property. At encryption side, the public key is generated. The secret key which is generated from this public key is the decryption key and finally the fragments received are decrypted separately at receiver end and integrated into the original data.

C) RSA Cryptosystem: RSA cryptosystem is one of the oldest method in asymmetric encryption and is widely used among the existing cryptosystem now-a-days. RSA cryptosystem was invented by three scholars named Ron Rivest, Adi Shamir, and Len Adleman. Hence, they named this cryptosystem with their names as RSA cryptosystem. RSA Cryptosystem algorithm is generally used in public-key cryptography and not private key crptofra. In this cryptography two keys are used to encrypt and decrypt the data in cloud namely a public key and a private key. The public key is used for encrypting messages and this key is general hence known to everyone. The encrypted messages are then decrypted by using the private key and hence he data is secured. During the verification process, the public key authentication is done by the server by using signature and a unique message along with its private key, which is called as digital signature. The signature is then returned to the client for verification at the server using the servers public key [12].

D) Identity Based Encryption: In identity based encryption (IBE), we have two keys public key and the private key. Unique information about the identity of the user is used to generate a public key. It allows to generate a public key from a known identity value for any party. A trusted third party known as Private Key Generator, creates the corresponding private keys [15] to reduces the complexity of the encryption process for both users and administrators. One of the best real time application for this type of encryption is Email Encryption.

3) Hash function

A) MD5- (Message-Digest algorithm 5): MD5 is the most commonly used hash algorithm in cryptography with 128-bit fixed-length hash value and output is of 128 bits. First the input message is segmented into block of 512 bit blocks and then this message is padded in order to get its total length to be divisible by 512. In MD5,[12] the data is encrypted using public key and decryption of the message using private key.

B) SHA: Secure Hash Algorithm (SHA) is a function that transforms a message of arbitrary length into a fixed length [10]. SHA computations for the messages encryption is calculated based on preprocessing and hash. In preprocessing method, pad message with zeros and the message length data, so that the message length to be a multiple of the

message block size. Finally parse the padded message to the message block size. In hash computation, we need to create a hash value by expanding a word first from a message block and then performing round operation on that word which is created by expanding the word. The SHA-512/224 and SHA-512/256 are same but with different initial hash values.

IV. RESULTS AND DISCUSSION

Table 1. Comparing Security Aspects Of Different Algorithms

Ref	Model	Algorithm	Advantages	Application	Drawbacks
[13]	public	Group sharing method using DH algorithm and key tree	Scalable, Efficient	Data storage in public cloud	Medium security
[14]	Public	Attribute based encryption, out sourced decryption	can run on top of any ABE schemes with outsourced decryption capabilities	cloud storage	Medium security
[8]	Private	data security and privacy	remote access, cost effective data storage	banking and finance service	Chance of threat
[9]	Community	Open source distributed file systems(TahoeLAFS and Xtreem FS)	fault tolerance, storage performance,	less storage nodes, less files for real data and usage	-
[7]	Hybrid	SNOWBIRDS Algorithm	Minimizes cost and processing time of satellite data	satellite image processing (e.g., weather hazards)	Specific for wind
[11]	Public	GLEnc	highly secured, data size minimized,	high security of public cloud records in the network	-

Table 2. Comparing Cloud Deployment Models

Deployment Model	Scalability	Security	Performance	Reliability	Cost
Private	limited	very high	very high	very high	high
Community	limited	high	very high	very high	medium
Public	very high	moderate	low to medium	medium	low
Hybrid	Very high	High	High	Medium to high	Medium

V. CONCLUSION AND FUTURE SCOPE

In this paper, algorithms for the deployment models has been compared to secure data in cloud. Algorithms for symmetric, non-symmetric, hash function has been discussed in TABLE I. TABLE II is comparison of different functional parameters of data security in cloud deployment models. Data security has advantages and disadvantages, the disadvantages discussed in the paper can overcome in future. Future enhancement can be done on public deployment model.

ACKNOWLEDGMENT

This survey paper has been supported by *Gopal K Shyam* [Reva University, Bangalore] who provided his insight and expertise that greatly assisted for this survey paper.

REFERENCES

[1] Yara AlHumaidan, Lama AlAjmi, Moudhi Aljamea, Maqsood Mahmud, "*Analysis of cloud computing security in perspective of saudi arabia*", published in IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), **2018**.

[2] T.Ramaporkalai, "*Comparitive study of security algorithms in cloud computing*", pubished in International Journal of Computer Engineering and Applications, volume **XII**, Issue **I**, pp.**124-121**, **Jan 2018**.

[3] Bih-Hwang Lee, Ervin Kusuma Dewi, Muhammad Farid Wajdi, "*Data Security in Cloud Computing Using AES Under HEROKU Cloud*", published in the 27th wireless and optical communications conference (WOCC2018), **2018.**

[4] Ahmed EL-Yahyaoui, Mohamed Dafir Ech-Chrif ELKettani, "*Data Privacy in Cloud Computing*", published in 4th International Conference on Computer and Technology Applications, pp.**25-28**, **2018**.

[5] M. Thangapandiyan, P. M. Rubesh Anand and K. Sakthidasan @ Sankaran, "*Enhanced Cloud Security Implementation using Modified ECC Algorithm*", published in International Conference on Communication and Signal Processing, pp.**1019-1022**, **India, April 3-5, 2018**.

[6] M. Thangapandiyan, P. M. Rubesh Anand, and K. Sakthidasan @ Sankaran, "*Quantum Key Distribution and Cryptography Mechanisms for Cloud Data Security*", published in International Conference on Communication and Signal Processing, pp.**1030-1035**, **India, April 35, 2018.**

[7] Remi Sahl, Paco Dupont, Christophe Messager, Marc Honnorat, Tran Vu La, "*High-resolution ocean winds: Hybrid-cloud infrastructure for satellite imagery processing*", published in IEEE 11th International Conference on Cloud Computing, pp.**883-886**, **2018.**

[8] Abhishek Mahalle, Jianming Yong Xiaohui, Tao Jun Shen, "*Data Privacy and System Security for Banking and Financial Services*

Industry based on Cloud Computing Infrastructure", IEEE 22nd International Conference on Computer Supported Cooperative Work in Design, **2018**.

[9] Mennan Selimi, Felix Freitag, Roger Pueyo Centelles, Agust Moll, "*Distributed Storage and Service Discovery for Heterogeneous Community Network Clouds*", IEEE/ACM 7th International Conference on Utility and Cloud Computing, **2014.**

[10] Sang-Hyun Lee, Kyung-Wook Shin, "*An Efficient Implementation of SHA processor Including Three Hash Algorithms(SHA-512, SHA512/224, SHA-512/256)*", **2018.**

[11] Lalu P, George, Dr.D.I.George Amalarethinam Bursar, Dr.Anjana S.Chandran, "*GLEnc Algorithm to Secure Data in Public Cloud Environment*", **2018**.

[12] Rishav Chatterjee, Sharmistha Roy, "*Cryptography in Cloud Computing: A Basic Approach to Ensure Security in Cloud*", **2017.**

[13] Celia Li, Cungang Yang, "A Novice Group Sharing Method for Public Cloud", published in IEEE 11th International Conference on Cloud Computing, pp**.966-969, 2018**.

[14] Changhee Hahn,Hyunsoo Kwon, Junbeom Hur, "*Toward Trustworthy Delegation: Verifiable Outsourced Decryption with Temper-Resistance in Pubic Cloud Storage*", published in IEEE 11th International Conference on Cloud Computing, pp.**920-923, 2018.**

[15] Deepanshi Nanda, Sonia Sharma, "*Security in Cloud Computing using Cryptographic Techniques*", published in International Journal of Computer science and technology, Vol.**8**, Issue **2**, pp.**66-69**, **April - June 2017.**

[16] Gary c. Kersler, "*An Overview of Cryptography*", https://www.garykersler.net/library/crypto.html, **31 march 2019**.

