Manasa V, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 10, Issue 7, (Series-V) July 2020, pp. 24-29

RESEARCH ARTICLE OPEN ACCESS

Survey on Cloud Computing: A Concept and its Data confidentiality

Manasa V*,Dr. Jayaram M N**
**Post Graduate Scholar, Dept. of ECE, JSS STU,Mysuru*
***Associate Professor, Dept. of ECE, JSS STU,Mysuru*

ABSTRACT
Cloud computing in plain words is the use of servers located remotely through the internet to save, monitor, and handle the data instead of using the server located locally or any personal workstation. The top reason for using the cloud is providing data access from anywhere. The public cloud computing in global market is projected to hit $330 billion, and about one-third of the budget of a company goes to cloud services, according to recent figures. But the main setbacks to cloud adoption are privacy and security. Some of the cloud security elements are data Integrity, confidentiality, availability, and privacy. This paper presents the extensive survey on the data confidentiality in the cloud, its issues and discusses some of the encryption techniques used and obfuscation techniques to make data more difficult to interpret which ensures the data confidentiality in the cloud, and also prevents unauthorized data access thus, ensuring that only the person who has the permission can access the data**.**

Keywords:cloud computing, public cloud, data confidentiality, encryption, and obfuscation techniques**.**

Date of Submission: 06-07-2020 Date of Acceptance: 21-07-2020

I. INTRODUCTION:

The widespread use of computers and Internet nowadays have introduced some genuine needs, such as- Larger Processing Power, More Storage, everything was online,better networking is needed, etcdue to all these constraints in terms of space,power,time and scalability, we got what we today knew as cloud computing. In recent years, cloud computing has gained a good reputation due to its, scalability, flexibility, self-service capability, shared computing resources, and affordability. One of the well-knowndefinitions of cloud computing is "In cloud computing without user's direct management, the cloud services like data storage, data processing power, etc and all of its resources are available at any time". The main characteristic feature of cloud includes multitenancy, more economical because of Pay as you go policy, Easy Maintenance, data availabilitybyproviding access to all thick and thin client platform through some standard techniques, On-Demand Self-Service, Cloud computing capacity to extend or minimize resources according to the customer-specific service needs which is its scalability and elasticity property[1],etc.Although there are so many advantages, sharing resources over the internet often entails risk, which raisesthe question about security and privacy of the data in the cloud [1].

Security elements in the cloud are- Data Integrity, confidentiality, availability, and privacy.

Data Integrity: Data integrity generally means ensuring the security of data from unauthorized third-party access for modification or deletion of data [3]

Data confidentiality: Confidentiality of data is about securing data from accidental, unauthorized, or illegal access, disclosure, or fraud. To ensure confidentiality of the data, authentication and access control methods are used

Data availability: The availability ensures the secure and timely access to cloud data or cloud computing services.

Data privacy: Data privacy concerns the laws, legislation, and measures that businesses take to protect the information they trust their consumers to.

Two data states usually risk their protection in cloud they are the data which is already stored in the cloud refers to cloud data, or any data that can be accessed through the Internet and the data going from and into the cloud[2], usually, the data going from and into is more vulnerable to risks than data stored in the cloud, because it has to move from one place to another.Some of the Security threats of cloud

computing are Data Breaches,hijacking of accounts,lack of Identity,controlled Access, Insecure Interfaces, Shared Technology susceptibility, information Loss, Deny of Service [4].A Service Level Agreement (SLA) which is the legal agreement in both the cloud service providers (CSP) and the users is being used to avoid such security issues[5].SLA will be responsible for providing all the cloud resources, services, and maintenance of all the issues, etc. In this paper, we discuss the two types of data that is the data moving from and intothe cloud and the data within the cloud which are usually at risk and discuss some of the encryption techniques used to encrypt the data in the cloud and obfuscation techniques used by which the data becomes very difficult to understand these ensuresthe confidentiality of the data in the cloud.

II. DATA CONFIDENTIALITY PRESERVING TECHNIQUES

A. An Auditing schemes.
B. A new framework using authentication and encryption techniques
C. A method to handle the encrypted data in the cloud with Distributed access storage.
D. An Encryption and Obfuscate technique.
E. A Classification technique using the K-NN algorithm.
F. A Classification technique using the Bagging and Boosting algorithm.
G. Shamir's key-based confidentiality scheme.
H. A computation over the cloud using randomization and shuffling technique.
I. Middleware Management Control software with an RC6 Encryption algorithm.

A. An Auditing scheme.

A confidentiality preserving technique called the auditing scheme is introduced using third-party auditor (TPA) in the Cloud Computing Environment (CCE). It consists of three entities Service Provider (SP) which is liable for storing user data and maintaining it,User who needs to save its data on Cloud, and TPA is having information to check theintegrityand preserving theconfidentialityof data saved in Cloud.Firstlythe user generates the secrete key , private and public key pair, and sends the public key to SP with file encoded using its private key. SP decodes the file with the usage of a public key and stores the file in Cloud Server. The user using HMAC-MD5 finds the metadata for the file and forwardsit to the TPA with the secret key for verification purposes.

Firstly, the TPA produces its own private and public key pair. The private key to perform a digital signature on the secret key and sends the digital signature to SP with its public key. SPverifies the signature so that the TPA is authenticated and then gets the metadata using the same hashing algorithm and sends it as a reply to TPA [6]. TPA checks for matching of the metadata sent by both the customer and the service provider. If both data are matched, then integrity and confidentiality are retained. TPA can use this result in the form of a report and can send that to the user.

B. A Framework using authentication and encryption techniques

A framework is proposed to maintain the confidentiality of data that is stored in the cloud using AES algorithm for encryption, SHA-1 for integrity, and Station-to-Station Key Agreement(STS)protocol for the authentication, For computing purpose many servers are used and every server which is present in a cloud network does the identical operation and are connected in ring form,Everyseveractslike a storage server and also an authentication server. Operation includes first the key is exchanged between the user and the cloud servers using the STS protocol to provide the authentication, now that server becomes an authenticated server. when uploadingthe file user encrypts the data with his private key and sends it to the authentication server. Authentication Server decrypts the file with its key and computes the SHA-1 algorithm for the original file that results inthe digest and eventually encrypts the data digest with its key using AES and sends the encrypted data to the recipient. Now the user has access to the storage server and likewise uploads the encrypted file to cloud storage, for downloading the data file the user can have a direct request to the system which is forwarded to the authentication server which then decrypts the file using the key[7]. It performs the SHA-1 on the file and matches it with the digest stored if itis identical then the file is validated and the file is sent back to the user, rather an error message is sent to the user indicating the corruption or the alteration of the data file

C. A method to handle the encrypted data in the cloud with Distributed access storage.

A mechanism to remove the risk of gaining unauthorized access to data by separating users based on their right of access, which guarantees the confidentiality of the data is introduced. According to which all the operations on data in the cloud should be performed with the encrypted format. To handle the encrypted data, the introduced model includes five elements as shown in Fig1.

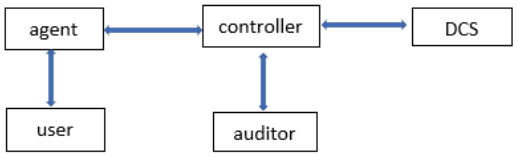

Fig1: User accessing the DCS

The agent is configured on the client's infrastructure and includes all processes to store, read, remove, encrypt, and decrypt data. The controller is liable for managing information about files in a database, authentication, and authorization procedure using the auditor [8]. DCS is a place where the encrypted data is fragmented and stored. To save the data in it the user gives theID information and raw data to the agent that is sent to the controller.The controller checks the access rights of the userusingthe auditor. After the user is successfully allowed to have direct access to the DCS nodes, the controller collects tokens from the DCS node which allows the agent to momentarily access the DCS node for data storage. Further, the data is divided into many parts and each part of the data is encrypted and transmitted which distributes the encrypted data between the nodes of DCS, and the agent gets data location information. Similarly, for getting the data from a DCS, a client sends a request for data after the successful verification and validation of it, then requests are sent to DCS nodes and the controller sends the token to the agent responsible for starting the loading fragments process. The agent decrypts and saves the data when the fragments are received and sends it to the user. The "Deniable Encryption" is used to encrypt the data.

D. An Encryption and Obfuscate technique.

A Cryptographic Technique (CT) to improvethe confidentiality of numerical and nonnumerical data in the cloud through both the encryption and obfuscate technique is proposed which is named Enc and Obfus_CT. The encryption is used to encrypt the non-numerical data and obfuscation technique which makes the data more difficult to interpretis used for the numerical data. During the process, both the CTs invokes the key management service (KMaas) to generate the key required for encryption and obfuscation that is sent to the user directly where the CSP is not involved in it. Once these techniques are done, both the data is uploaded into the cloud, key storage components of Key management service is used to store the key. It is found that Enc_CT and Obfus_CT produce 90 percent and 91 percent once after the simulation, the Enc and Obfus_CT produce 93 percent of the security level[9]. Hence, the Enc and Obfus_CT scheme produces a maximum-security level and provides confidentiality forboth non-numerical and numerical data stored in the public cloud environment.

E. A Classification technique using the K-NN algorithm.

Confidentiality is provided using the data classification model for cloud computing. The K-Nearest Neighbour (K-NN) is a machine learning technique used in a designed simulation background to classify the data based on their security needs[10]. The data was divided into two categories, confidential and non-confidential. In K-NN Algorithm, it first determines the set of n identified samples and the value of K set to one, then calculates the distance between the new input data and all the trained data and determines the K-nearest neighbor based on the K-th shortest distanceand finally determines the class of the new input based on a majority vote which may result in two classes the confidential data and the non-confidential data. Data that is classified as confidential maycontain Personal Data, Business material, etc, non-confidential data may be Public data that includes press announcements, marketing material, etc. The RSA encryption algorithm is used to encrypt the confidential data once after the classification is completed to protect its access from unauthorized users. The Virtual Machine (VM) directly receives the public data without encryption which further processes the data and interacts with the server for data storage.

F. A Classification technique using the Bagging and Boosting algorithm.

Another data classification approach is proposed in whichit identifies the information into confidential and non-confidential usingenhanced bagging and boosting methods. Upon categorization of the information, the blowfish algorithm is implemented to protect the confidential data and non-confidential data is being sent to the cloud lacking encryption.The Bagging technique is focused on the voting system and this uses an enhancer where the training set of data is given, a training set as input data, then manipulating the enhancer multiple times by altering the dispersion of items in the training set.Generated classifiers were combined to produce a final classifier that, in effect, categorizes the given input data.The AdaBoost method is used in the Boosting technique which generates a group of classifiers and conducts weight-based voting. The weight of each classifier in this strategy is

considered in producing the output. Blowfish algorithm commonly used for the encryption of the data which loops by almost16 times a single event.The enhanced bagging and boosting technique achieves improved results compared to the K-NN algorithm[11] and thus reduces the time required forcategorization and reinforces accuracy and solves our confidentiality concern in a cloud background, as well as dividing data into sections and separately stored in different cloud partitions to enhance safety. This type of classification scheme will reduce the overhead and time for securing the entire data.

G. Shamir's key-basedconfidentiality scheme.

A Shamir's Key system provides Confidentiality based on the Distribution of secret keys (SKDC) by residing the cloud data with polynomial interpolation.This secret key sharing scheme is used when the multiple users want to store some information securely in which the polynomial equation is created which consists of the secret as the first coefficient and the remaining variables in the equation at picked at random that improves the security level in the cloud infrastructure.

The different curves for the given user information using the polynomial interpolation are used which includes the user input say Si=S1, S2,….Sm with the information points Pa=P1,P2….Pb. According to polynomial property Si=S1*[P1] +S2*[P2]+ S3*[P3]+…..+ Sm*[Pm] by using the lookup table on information point storage from Shamir key distribution

K=Si mod n where $1 \leq i \leq m$

The $1 \leq i \leq m$ in the range of the cloud user information obtained and K is the shared secret key. The 'K' is the key secretly held by users of the public cloud to increase confidentiality.

To provide confidentiality in the higher rate this scheme allows the data from multiple users [12]. Therefore, to reconstruct the secret the sharing of the information from the entire user is required.In this way,a new key distribution scheme is used to achieve a higher percentage of confidentiality.

H. A Computation over the cloud using randomization and shuffling technique.

A method is proposed were the cloud users will get more control to ensure the confidentiality of numerical data. For a secure matrix multiplication over cloud network using column-row shuffling, randomization, and size alteration of matrices[13]. In this method, two protocols are proposed one to hide the data and the other to retrieve the data. Here one client and two servers are used to demonstrate two protocols where for one server a randomized unencrypted data is sent and for another server, the false unencrypted randomized data is sent.

Steps to hide the data :

1.Assume A and B are two matrices with a size of (mxn) and (nxk).

2. A random square matrix D which is a shuffled unit matrix (nxn).

3. Compute A1 = A x D and B1 = D−1xB, where D and D−1 willbe nullified.

4.The matrix I(mxz) is generated, where z is a natural number, and z = nx0.3,The columns of Iare randomly inserted into A1.Thus, the matrix is G of size (mx(n+z)) is produced as a result.Similarly, another matrix J(zxk) is used and the matrix is H of size ((n + z)xk) is produced.where the G and H are forwarded to one server.

5.The values of I and J are hidden by usinga random diagonal square matrix K of size and calculate K−1, compute M = IxK, and N = K−1xJ.

6. Random diagonal square matrices P(mxm) and Q(kxk) are generated and

R = PxM; and S = NxQ. Send R and S to another server.

Steps to retrieve the data

The first server computes C-1 = GxH of size mxk. The second server computes T = RxS. The client now performs the simple computation C= C-1 –(P-1xTxQ-1) to get the actual result C from C-1 this operation is only known by the client. The main contribution in this strategy involves altering matrix size along with shuffling and randomization to mask the actual data transmitted.

I. Middleware Management Control Software With The RC6 Encryption Algorithm.

A new data storage scheme that ensures data confidentiality with time-efficiency is introduced. A Middleware Management Control (MWM) software is developed between customer and cloud providers as an intermediary software layer[14]. MWM rotates each row of the file uploaded by the consumer to columns then vertically divides it into subfiles according to columns even /odd. Then divide into further even and even, even and odd, odd and even, odd and odd subfiles and all the sub-files are encrypted using the RC6 algorithm which ensures the confidentiality of data.

Finally, different cloud space is chosen to upload the encrypted sub-file, and it is mapped in the location table. If

the user requests a specific file then this software downloads each sub-file, decrypts it, and assembles them into one file based on the location table
and changes rows to columns and returns them to t he user. Hence confidentiality is provided and by dividing data into

sub-documents and then applying the RC6 encryption algorithm to each.

III. COMPARATIVE ANALYSIS:

A comparative analysis is a methodology of comparing two or more things with the idea to come upon something new about one or all the things being compared. Here the comparative analysis of all the schemes discussed is done based on some parameters such as algorithm used, confidentiality, integrity, authentication provided by each scheme, attacks that can occur, its advantages, and limitations.

Table 1: A comparative analysis of the encryption and obfuscation techniques

Method	Algorithm	Key	Possible Attack	Advantage	Disadvantage
Auditing Scheme	Digital Signature, HMAC-MD5	Secret key, Asymmetric key	Brute Force Attack	Provides Integrity With confidentiality	Secret key sent in an insecure channel
A new framework	AES, SHA-1, STS protocol	Session Key and Symmetric key	Unknown Key Share Attack	AES used which is more Secure and faster	SHA-1 is considered as unsafe prone to a collision attack
Distributed access Storage	Deniable Encryption	Two or more key	Common pollution attack	Different parts of data stored in different space of cloud	User waiting time increases because of a token system used
Encryption and obfuscation	AES, DES	Multiple keys required for both the techniques	Replay and phishing attack	Both numerical and Nonnumerical Data is protected	Difficult to manage multiple keys
Classification Based	K-NN And RSA	Asymmetric key	Brute force attack	Only sensitive data is encrypted	KNN's efficiency Reduces as data grows
Classification based	Bagging, boosting, blowfish	Symmetric key 32-448 bits	Birthday attack	classification is more accurate than KNN	No authentication, Slower Decryption
Shamir's key distribution	Polynomial Interpolation	Distributed key	Man, in middle Attack	Each part of the secret has the same size as the secret	used when the group of members needs to maintain the secret
Computation Over Cloud	Randomizati on And Shuffling	No keys used		The only the client knows the actual way of Computation is done	Applicable only for numerical Data
Middleware Management Software	Rotation, Division, RC6 encryption	Symmetric key	Linear attack	A client only knows the actual way of Computation is done	Rotation, division becomes difficult for large size of data

IV. CONCLUSION:

Cloud computing with the feature of data availability, highly capable network, cost-effective computers, storage devices,

and the hardware made it as the best solution for data storage since resourcesit issince resources are sharing over the internet always it is prone to one or the other security threat this paper discusses one of the major security elements "Data Confidentiality", many of its issues and some of the encryption and obfuscation technique for preserving the data confidentiality. The encryption techniques such as AES,RSA,RC6,Blowfish with the classification methods, and the obfuscation technique like Randomisation, shuffling, division,rotation, etc are used. In future accurate classification methods with the fragmentation technique and then secure encryption algorithmcan be appliedto provide data confidentiality so that the classification techniques reduce the time complexity because only the required data is encrypted. After classification, the confidential data is fragmented into smaller segments of data and then a secure encryption algorithm is applied which can be Elliptic curve cryptography which has a shorter key size and usually used key agreement, digital signature, and another task.

REFERENCES

[1]. Jihad Qaddour, "Security Threat and Challenges Analysis of Cloud Computing with Some Solutions", International Journal of Computer Science and Telecommunications,Volume 9, Issue 7, December 2018

[2]. Madini O. Alassafi, Robert Walters, Ahmed Albugmi, Gary Wills, "Data Security in Cloud Computing", Fifth International Conference on Future Generation Communication Technologies (FGCT 2016), IEEE.

[3]. M. Subhashini, Dr. P. Srivaramangai, "A Study on Cloud Computing Securities and Algorithms", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Volume 3, Issue 3,March 2018.

[4]. Manivannan Chakkaravarthy, AnamikaBhattacharya, Deepak R Bharadwaj,"Cloud Threat Defense – a Threat Protection and Security Compliance Solution", IEEE 2018, International Conference on Cloud Computing in Emerging Markets.

[5]. Chandan Prakash, Surajit Dasgupta, " Cloud Computing Security Analysis: Challenges and Possible Solutions", IEEE, International Conference on Electrical, Electronics, and Optimization Techniques, 2016

[6]. Suneeta Mohanty,Prasant Kumar Pattnaik, Raghvendra Kumar, "Confidentiality Preserving Auditing for Cloud Computing Environment", IEEE, 978-1-5386-2599-6/18, 2018

[7]. Deepak Singh,Harsh K Verma, "A New Framework for Cloud Storage Confidentiality to Ensure Information Security", IEEE,2016 Symposium on Colossal Data Analysis and Networking (CDAN).

[8]. Andrey N. Rukavitsyn, Ivan I. Holod, Konstantin A. Borisenko, Andrey V. Shorov, " The Method of Ensuring Confidentiality and Integrity Data in Cloud Computing", IEEE, 2017.

[9]. S. Arul Oli, Dr. L. Arockiam, " Confidentiality Technique to Encrypt and Obfuscate Non-Numerical and Numerical Data to Enhance Security in Public Cloud Storage", World Congress on Computing and Communication Technologies, IEEE, 2017.

[10]. Munwar Ali Zardari, Low Tang Jung, Nordin Zakaria, " K-NN Classifier for Data Confidentiality in Cloud Computing",IEEE,2014.

[11]. Rasmeet Kour, Suparti Koul, Manpreet kour, "A Classification Based Approach For Data Confidentiality in Cloud Environment", IEEE 2017 International Conference on Next Generation Computing and Information Systems.

[12]. M. Sugumaran,D. Kamalraj, B. Balamurugan, S.Jegadeeswari, "Shamir's Key-based Confidentiality on Cloud Data Storage", IEEE, 2015.

[13]. Mahboob Shaheen, Khaled M. Khan, "Empowering Users of Cloud Computing on Data Confidentiality", 3rd International Conference on Cloud Networking, IEEE 2014

[14]. Karim Timraz, Tamer Fatayer, Tawfiq Barhoom, "A Confidentiality Scheme for Storing Encrypted Data through Cloud", IEEE 2019.