Security and Privacy Challenges in Cloud: Survey and Research Directions

*V. Swathi1, Dr. M. P.Vani2

1Research Scholar, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Vellore-Tamil Nadu, India.
2Associate Professor-SITE VIT University, Vellore- Tamil Nadu, India
Corresponding Author: V. Swathi

ABSTRACT

Cloud Computing is known to be a new computing model which provides reliable, secure and quality assured computational environment for users. Most of the users prefer to store their data inside the cloud in an encrypted format to decrease the security concerns. However, to perform any operation on data at server, cloud needs to first decrypt the data. This operation might cause the challenging issues e.g. confidentiality along with privacy of confidential data, stored inside the cloud. Therefore secure outsourcing mechanisms are much needed to protect the sensitive information by enabling computations with encrypted data as well as to verify and validate the computation end result. To design such type of mechanisms is a big challenge. This paper presents an extensive literature study and review of latest advances, developments and new methodologies in cloud computing, concerning security and privacy challenges. We have discussed various design challenges and some significant characteristics of these schemes. Finally, we have outlined some open problems in this domain and our further research aims and directions.

Keywords: Cloud, Security and privacy, Data integrity, Complexity, Biometrics, Identification.

I. INTRODUCTION

Most of the users prefer to store their data inside the cloud in an encrypted/unoriginal format to decrease the security concerns as how to protect the data that is processed and generated by the customers, is becoming the major concern [16][18][22] in the present day situation. Basic advantage of cloud computing [19][20][21] is that it is having the benefits of centralized large computational power, space and efficiency, so that the clients can outsource their complex problem to the cloud for computation purpose. Although the cloud computing is being used to outsource large-scale computations to the cloud, data privacy [36][37] has become a major issue. Though, to perform any operation on data at server, cloud needs to first decrypt the data. This operation might cause the challenging issues [34][35] like - confidentiality along with privacy of confidential data, stored inside the cloud. Cloud computing is a computational mechanism, which is used for the convenient non-demand network access to the shared pool of the computing resources which is having the greater efficiency as well as large computational power [23][24][32][33]. General framework of cloud-client model is shown as below diagram:-

![Figure 1: Secure outsourcing in cloud computing](image-url)
Cloud is having the great potential of robust computational power to the aggregated management of the elastic resources. The outsourced problem constraints from the customers’ side may contain private and sensitive information e.g. Personal identifiable business information, sensitive research data etc. So to protect this data from unauthorized use, customers need to encrypt their data prior to outsourcing, but further performing the computations on this encrypted data makes a very hard problem for the cloud server [25][26][27][28].

A. Motivation to the problem
There exist a much powerful thrust to provide the security at various infrastructure levels inside cloud [29][30][31], while any outsourced computing or any third party computations are being performed. To resist against unauthorized information leak, sensitive data essentially has to be encrypted before outsourcing to cloud. Common data encryption methods [38][39] in essence limit cloud from performing any meaningful operation of the underlying plaintext data, causing the computation over encrypted data a very difficult problem. Homomorphic encryption [6][7][40] is evolved as an special form of encryption that use to allow the computations to be carried out on ciphertext, therefore generating an encrypted result which, when decrypted, will give a match with the result of operations, performed on plaintext data. In 2010, Gentry proposed the Fully Homomorphic Encryption mechanism, which was a great breakthrough in cryptography [41][42].

B. Organization order of the paper
In rest of the paper, section 2 discusses some security and privacy challenges in this domain. An extensive literature review is presented in section 3. Section 4 summarizes some significant issues of cloud storage and need for the data robustness schemes. Our further research directions are given in section 5. Finally section 6 concludes the paper.

II. SECURITY AND PRIVACY CHALLENGES
Many traditional and modern techniques like partitioning, virtualization and other cloud based applications are being utilized in cloud scenarios.

Figure 2: FHE application in cloud computing

Fully Homomorphic encryption (FHE) is a tremendous way to perform processing and computations on the encrypted data in cloud environment, which is being used by any third parties without the knowledge of private secret key.

B. Organization order of the paper
In rest of the paper, section 2 discusses some security and privacy challenges in this domain. An extensive literature review is presented in section 3. Section 4 summarizes some significant issues of cloud storage and need for the data robustness schemes. Our further research directions are given in section 5. Finally section 6 concludes the paper.

II. SECURITY AND PRIVACY CHALLENGES
Many traditional and modern techniques like partitioning, virtualization and other cloud based applications are being utilized in cloud scenarios.

Figure 3: Overview of privacy challenges
We are going to discuss several security and privacy challenges in this section.
A. Offloading and Partitioning security challenges
During the offloading procedure [35][36][39][40], it need to access the cloud through wireless networks. Since the mobile users do not have any access as well as control over their offloading processes, so, there involved a risk of unofficial access to offloaded content. The confidentiality and integrity challenge grows because after the execution of offloaded content, if result is modified, the mobile devices cannot perform verification easily about the correctness of results [70].

B. Data security challenges
The data related challenges include data loss, data recovery, data locality as well as data privacy. The data loss and data breach break mainly the two security requirements such as integrity and confidentiality. Here, the data loss means – the users’ data is in error condition that damaged or skipped by any physical means during processing, transmitting or storage. The correctness of data becoming one of the concerns for mobile users in cloud scenario [71][72].

C. Security challenges in Virtualization aspect
In cloud end, an image of virtual machine (VM) of the mobile device is pre-equipped and the tasks of the mobile device are offloaded to the virtual machine for processing [36][39][73]. This virtual machine is also called thin virtual machine or phone clone. The prime function for virtualization is to give various virtual machines running on same mobile devices but isolated to each other.

D. Cloud applications security challenges for mobility
Cloud based mobile application level attacks can affect the integrity and confidentiality of both the data and applications by several procedures e.g. integrating malwares. Malwares like worm, rootkit, botnet etc. are unfavorable, intrusive applications or programmed codes. The targets of these malwares are to run with intentions at mobile devices or attach with applications without users’ compliances. As a consequence, the functionalities of mobile applications can be changed.

E. Security challenges for mobile devices
Although there are password or pattern based locked features; many mobile users do not use these features. And the identity module card inside the mobile device also can be taken aside from device and accessed by unauthorized persons. Moreover, most of mobile devices are lack of security mechanism against threats. The attackers can attack by utilizing different availability attack techniques such as by sending high malicious traffic stream, huge messages to targeting mobile devices to make unused or reducing the capability [48]. The battery power exhaustion attack is another kind of availability attack where after attacking, the mobile devices start to discharge its battery power rapidly. This attack is unique for mobile device as it is performed by utilizing vulnerabilities of wireless networks, and mobile users are unaware about this kind of attack. The functionalities of present mobile platforms [74][75] are quite close to personal computers but with extra features and these platforms for mobile devices support many applications. Hence, to ensure confidentiality and integrity of these applications, there need to secure the mobile platforms also.

There are three kinds of storage available in mobile devices such as on device storage, plugged in storage and identity module storage. Generally, user’s personal data, applications and others are stored in these storages. But if a mobile user avail cloud services, its data and applications are copied in the cloud storages [76]. So, if the mobile device is stolen or lost, then it becomes an important point such as the attackers can get access to the mobile devices and access to the cloud as well.

F. Privacy challenges in cloud
Cloud storage and processing in multiple locations raise privacy problems. The cloud servers of service providers are located at different regions and countries. For example, Google’s cloud servers are located almost around the world such as seven locations in Americas, two locations in Asia and three locations in Europe. Moreover, it is important to users to get cloud hosting location’s information as the law differs from one country to another. Several mobile applications are available which may be unsafe due to having hideous functions, collecting unintentionally users personal information e.g. hobbies, locations, and may spread illegally. The unwanted emails or junk emails can violate the users’ privacy [47][77].

The context awareness, enabled by the sensors on mobile devices, is one of the main features of mobile applications that differ from personal computers. The context provides information to service providers by giving users’ context and hence, the service providers can provide services with respect to the requirements of users. These location-aware applications and services specifically raise privacy concerns for mobile devices. These can either be user invoked or service provider invoked, and need the user location’s knowledge to deliver the location based services.
III. LITERATURE REVIEW

This section extensively represents the research work and developments that has taken place in past years. Cong Wang et. al. [1] has described about secure outsourcing mechanism which is used to solve large scale systems of linear equation of cloud. Applications of Gaussian elimination and LU decomposition approaches are becoming expensive due to large scale LE problems. Hence, another approach named iterative method is used [16][17], [2][43][44] presented the protocol as secure computation protocol, is used for bridgeing secure storage also secure computation auditing in cloud. [3][45] Has mainly investigated about secure outsourcing of widely applicable linear programming computations. Present mechanism explicitly decomposes LP computation outsourcing [46][47] into the public LP solvers which are running on the cloud also those private LP parameters owned by the customer. [4][48] has mainly described practical outsourcing schemes, in which they described how to securely outsource modular exponentiation, that presents the computational bottleneck in public key cryptography [49][50] on computationally limited devices. [5] Discussed about problem of simultaneously achieving scalability, fine grainedness and data confidentiality of access control. Here, they have defined access policies which are based on data attributes; also they delegate the computation tasks to the untrusted servers without changing any underlying contents in the data. Craig Gentry et. al. [6][51] proposed FHE scheme means schemes that allows to evaluate circuits over the encrypted data which is unable to decrypt. Final solution of this complete issue comes under 3 steps. (i) They have provided general result, which is used to construct the encryption scheme. (ii) Later described public key encryption scheme using the ideal lattices which is almost bootstrapable. (iii) They showed modification of scheme to reduce depth of decryption circuit. An application scenario of FHE is given in fig 4 below –

![Application Scenario: FHE](image)

[7] Has mainly focused on storing of data on cloud in the encrypted format using FHE. Results are downloaded from client machine only. Hence, users data is never stored in plaintext on public cloud [52][53]. [8][54] Described about Cloud Computing Algebra Homomorphic Encryption Scheme Based on Fermat’s Little Theorem. They have proposed this application on cloud computing for efficient security. [9][55][56] listed out three fundamental questions about FHE, purpose of FHE, state of FHE. Also given clarity on different terminologies which are in use and prove connections between FHE notions. [10][57] described about Data Banks and Privacy Homomorphism. In this, it appears similar to that encryption functions which permits encrypted data [58][59] should be operated on without prior decryption. [11] proposed hybrid homomorphic system which is applied to the banking data and is used to perform the operations on encrypted data avoiding decrypting. Khalid El Makkaoui et. al. [12][60] has proposed Challenges of Using Homomorphic Encryption to Secure Cloud Computing. Mainly they goes through challenges facing HE methods for allowing the suppliers of cloud to perform operations on encrypted data, also provides same results after completion as if they are performing these calculations on raw data. Muhammad Baqer Mollah et. al. [13][61] discussed the data which is stored in cloud infrastructure and actual execution is shifted to cloud environment. [14] aims to enhance the understanding of security issues associated with the cloud storage, also highlights the importance of data integrity schemes for the outsources data, and have presented taxonomy of existing data schemes used for the cloud storage. Changhee Hahn et. al. [15] proposed an efficient and privacy preserving biometric identification scheme in cloud computing. This scheme exploits a symmetric homomorphic encryption algorithm in biometric identification, thus achieving both security and efficiency. The client encrypts and enrolls his/her fingerprint. For identification, the client encrypts and sends a newly scanned fingerprint to the cloud. The previous key used to enroll is not re-used but a fresh key [63][63] is generated at every identification to encrypt the fingerprint. In the proposed scheme, Initial client enrollment phase occurs in which the first client has FingerCode $m_i = [x_{i1}, ..., x_{iN}]$ and a random key $k_i = [k_{i1}, ..., k_{iN}]$. 

![Application Scenario: FHE](image)
he encrypts $m_1$ using $k_1$. The client then sends the encrypted file to the server, which re-encrypts the file, then sends to the cloud. Next, the client, server, and cloud establish a secure channel among them to run a key-exchange protocol. Further second and subsequent clients enrollment is done. By iterating this process for $n$ number of clients, the cloud has encrypted FingerCodes as follows:

$$\sum_{i=1}^{n} Enc_{k_1+k_a+m_1}(m_1)....Enc_{k_1+k_a+m_n}(m_n)$$

the encrypted result is decrypted and cloud computes the Euclidean distance between FingerCodes and sends it to the server to verify whether the client is legitimate by using some threshold i.e. if $\text{dist}_{\text{min}} \leq t$ (where, $t$ is threshold), the client is successfully identified. Cloud computing services [23][64] are organized as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) [65][66]. Under the umbrella of these services a number of new services have also emerged, e.g. Data Integrity as a Service (DIaaS), Database as a Service, Logging as a Service, Security as a Service etc [67][68][69]. was introduced with or without third party auditor (TPA). Privacy issues such as leakage of data and user anonymity were identified due to the introduction of TPA. Thus, privacy preserving data integrity schemes were developed to overcome the privacy concerns of the users.

IV. ISSUES OF CLOUD STORAGE AND NEED FOR DATA ROBUSTNESS SCHEMES

This section discusses the significant characteristics for data robustness schemes along with design challenges.

A. Characteristics of data robustness schemes

All data integrity schemes have following significant properties:

1) Remote verification: Remote verification means that a verifier should not retrieve file blocks from the remote server for verification. In this verification, the cloud accesses block of a data file and generate a proof of possession, which is later verified by a client to validate data integrity.

2) Unrestricted challenge frequency: means that there should be no restrictions on the numbers of challenges made by a client or auditor to verify the integrity of remote data. A client executes the verification protocol with intervals to timely identify any data corruption. The frequency of challenge requests is directly affected by the computation efficiency of the data integrity scheme [78][79]. If verification process of a data integrity scheme is computationally expensive, then the client will use it less frequently and as a result unbounded challenge frequency would suffer.

3) Soundness: property of data integrity schemes ensures data reliability. The nature of a data integrity schemes is tamper evident, not tamper resistant. Therefore, if metadata are tampered with, or data get corrupted intentionally by the CSP [80] or unintentionally, this should be timely distinguished by a data integrity scheme. If the CSP can pass a challenge request without holding the data or with corrupted data, then a client will never be able to identify data corruption promptly, and the value of the data will be lost. Therefore, the reliability of data directly demands the soundness of a data integrity scheme for cloud storage [81][82].

4) Robustness: Data integrity schemes of deterministic nature have a limitation of computational efficiency and are not applicable for large datasets. To overcome these limitations, a probabilistic approach is adopted. Probabilistic data integrity schemes work on randomly chosen samples of data for integrity verification, which raises the issue of identification of small (bit or byte level) corruptions. Robustness property enhances the soundness of data integrity schemes by providing the identification of minor level corruptions in data.

5) Data recovery: For a data integrity scheme, it is not sufficient just to identify the misbehavior by the remote storage server. As cloud users are also interested in, complete data recovery to avoid an unpleasant situation. Data integrity schemes are divided into two categories depending on their tendency. Usually, error-correcting codes (ECC) are used by the schemes providing data recovery.

6) Dynamic data handling: Data can be either static i.e. backup or archival data or dynamic nature i.e. supporting operations like - insertion, deletion, and modification. Providing data integrity for dynamic data is more challenging than for static data or append-only data. Most of the schemes proposed in the literature are not capable to handle dynamic data.

7) Privacy preserving: Privacy concerns are introduced due to public verifiability. As the data owner will not allow the disclosure of his private data [83][84] to a third-party auditor, the privacy preservation property demands that a third party auditor should not obtain any confidential information about the users data but can still verify the integrity of outsourced data.
B. Design challenges

Some of the notable design challenges are listed below -

1) Computation and Storage efficiency: All data integrity schemes preprocess the data before outsourcing it to a cloud storage server, similarly, at the cloud storage server [41][85][86]; metadata are generated from original data. The overhead of performing such processing may affect the computation efficiency of data integrity schemes. Computation efficiency of preprocessing phase does not matter for small datasets, but it has a serious impact for large datasets. Computation cost on the server side for the proof generation limits on how frequently user can verify the integrity of outsourced data. Primitives used by the data integrity scheme as metadata also impact computation time. For integrity verification and data recovery by any data integrity scheme, additional metadata is required along with the original data. Furthermore, data de-duplication plays a significant role in the reduction of storage [87][88] though will incur some cost.

2) Communication efficiency: In the case of dynamic data, this communication overhead also includes the updates verification. Primitives used for metadata have an impact on the communication cost [89][90]. Three aspects describing the communication efficiency of data integrity schemes are (1) Overhead of the initial data transfer along with metadata, (2) challenge requests for possession verification by data owner [91][92] and (3) challenge response for proof of possession sent back by the cloud storage server [93].

3) Security challenge: Data integrity schemes are vulnerable to a variety of attacks. Therefore, care must be taken in designing such a scheme. Following are the possible attacks that may be launched against a data integrity scheme:-
  - Data deletion attack
  - Tag forgery attack
  - Replace attack
  - Replay attack
  - Data leakage attack

While designing a data integrity scheme, care must be taken so that the metadata used for verification may not result in leakage of original data during the verification protocol. Modern cryptographic techniques [6][7][94] such as ”Multi-Prover Zero-Knowledge Proof System” can provide protection against tag forgery attack and data leakage attack. Using homomorphic verifiable responses and hash index hierarchy ”Collaborative PDP” ensures that any adversary capable of eavesdropping the communication channel will not be able to extract any information from data exchanged in verification protocol [95][96].

4) Reduced disk I/O: Disk I/O efficiency of a data integrity scheme is derived by the overhead in block access and metadata access for proof generation on cloud storage server. Accessing all blocks for proof generation impacts the efficiency of a scheme and makes the scheme impractical for large datasets.

V. OPEN ISSUES AND RESEARCH DIRECTIONS

How to protect customers confidential data (e.g. Business financial records, personal research data etc.), which has to be prepared and generated during the computation is becoming the major security concern [21][97]. So to protect this data from unauthorized use, customers need to encrypt their data prior to outsourcing, but further performing the computations on this encrypted data makes a very hard problem for the cloud server [98][99]. Some open issues and our future research directions are discussed as follows:-

A. Problem Identification

Most of the users favor to store their data inside cloud environment [57][100] in an unoriginal form to decrease the security concerns [101]. However, to execute any operation on the data, which is residing at server, cloud needs to first decrypt the data. This operation might create the confidentiality and privacy issues of data stored in the cloud. Homomorphic encryption is a kind of encryption mechanism that give ability to users for computations to be prosecuted on cipher text itself, thus producing an unoriginal result when decrypted it shows similarity on the result of operations prosecuted on the plain text.
While on another point of view, the operations which are being performed may not transparent to customer, so customer needs to verify the result of the outsourced problem after the computation is performed. Most significantly, security is a prime concern that prevents the adoption of computation outsourcing in the cloud.

B. Future Research Directions
The aim of the research is to develop new practically secure and efficient schemes for the data security in cloud. This aim is fulfilled by using the following objectives -

- We will come up with optimal Homomorphic encryption (HE) schemes and apply it to secure computation in cloud. Homomorphic encryption is a way to perform processing and computations on the encrypted data in cloud environment, which is being used by any third parties without the knowledge of private secret key.
- We will also present the practical applicability and security analysis of the proposed schemes.
- One limitation of homomorphic encryption is that it does not support for multiple users. We will come up with the possible ways to solve this problem.
- In our implementation, we may work on a virtual platform as a Cloud server, a VPN network that links the Cloud with the customer and then simulate the different scenarios. We can connect to the AWS DynamoDB service through the Eclipse IDE. Further we can build our own AWS model for proposed system and will try to implement the proposed FHE scheme, and later perform simulation over it. OpenNebula, Eucalyptus types of open source tools can be utilized for virtual cloud environment experiments.

VI. CONCLUSIONS
Customers need to outsource their problem to the cloud server for computation in a secure manner that brings new challenges for customer’s data privacy and confidentiality. This paper started with the discussion of introduction to the entire problem domain, security and privacy challenges involved in it. Then we have presented an extensive review of work done in past years. We have mentioned the significant characteristics of robustness of schemes along with its design challenges. Further, we have listed some vital issues in this domain and our future research directions.

REFERENCES
Security and Privacy Challenges in Cloud: Survey and Research Directions


Security and Privacy Challenges in Cloud: Survey and Research Directions


