A Review on Data Security in Cloud Computing

Fakhruddin Noori
M-Tech “CSE”
Computer Science Faculty Rana University
Kabul Afghanistan.

Abdul Ghafar “Omerkhel”
Mcs
Computer Science Faculty Rana University,
Kabul Afghanistan

Abstract:- Cloud computing is an Internet-based computing and next stage in evolution of the internet. It has received significant attention in recent years but security issue is one of the major inhibitor in decreasing the growth of cloud computing. It essentially shifts the user data and application software to large datacenters i.e. cloud, which is remotely located, at which user does not have any control and the management of data may not be completely secure. However, this sole feature of the cloud computing introduce many security challenges which need to be resolved and understood clearly. One of the most important and leading is security issue that needs to be addressed. Data Security concerns arising because both user data and program are located in provider premises. In this study, an attempt is made to review the research in this field. The results of review are categorized on the basis of type of approach and the type of validation used to validate the approach.

Keywords:- Data security, data cloud concealment, cloud security, review

1. INTRODUCTION

Cloud computing is an emerging technology which recently has drawn significant attention from both industry and academia. It provides services over the internet, by using cloud computing user can utilize the online services of different software instead of purchasing or installing them on their own computers. According to the National Institute of Standard and Technology (NIST) definition, cloud computing can be defined as a paradigm for enabling useful, on-demand network access to a shared pool of configurable computing resources [1]. According to Gartner [2] cloud computing can be defined as a style of computing that delivered IT capabilities ‘as a service’ to end users through internet.

According to recent survey by International Data Group (IDG) enterprise, the top three challenges to implementing a successful cloud strategy in enterprise vary significantly between IT and line-of-business (LOB). For IT, concerns regarding security is (66%) and 42% of cloud-based projects are eventually brought back in-house, with security concerns (65%) [3]. A survey conducted by International Data Corporation (IDC) in 2011 declares that 47% IT executives were concerned about a security threats in cloud computing [4]. In survey conducted by Cisco’s Cloud Watch 2011 report for the U.K. (research conducted by loud house) 76% of respondents cited security and privacy a top obstacle to cloud adoption [5].

Data security is a major concern for users who want to use cloud computing. This technology needs proper security Principles and mechanisms to eliminate users concerns. Most of the cloud services users have concerns about their private data that it may be used for other purposes or sent to other cloud service providers [6]. The user data that need to be protected includes four parts [7] which are: (i) usage data; information collected from computer devices (ii) sensitive information; information on health, bank account etc. (iii) Personally identifiable information; information that could be used to identify the individual (iv) Unique device identities; information that might be uniquely traceable e.g. IP addresses, unique hardware identities etc.

The European Network and Information Security Agency (ENISA) identified thirty-five risks and these risks are divided into four categories: legal risk, policy and organizational risks, technical risks and risks that are not specific to cloud [8]. From these risks, the ENISA identified eight most important risks. Out of which five risks concerns directly or indirectly related to the data confidentiality. These risks include isolation failure, data protection, management interface compromise, insecure data deletion and malicious insider. Similarly, The Cloud Security Alliance (CSA) identifies the thirteen kind of risks related to the cloud computing [9]. Out of these thirteen risks CSA declares seven most important risks [10]. Five of these seven risks are directly or indirectly related to the data confidentiality which includes: account service, traffic hijacking, insecure application programming interfaces, data loss/leakage and malicious insiders.

Different countries, IT companies, and the relevant departments have carried out the research on cloud computing security technology to expand the security standards of cloud computing. Existing security technology reflected in six aspects[11,12] which include: data privacy protection, trusted access control, cloud resource access control, retrieve and process of cipher text, proof of existence and usability of data and trusted cloud computing. To enhance the data security the data can be converted into cipher text but this may cause to lose many features when data is converted into cipher text.

There are two widely used methods to retrieve the cipher text. First, there is a safety index-based approach which establishes a secure cipher text key words indexed by checking the existence of key words [13]. Second, there is a cipher text scanning-based approach which confirms the existence of key words by matching each word in cipher text [14]. [15] Lists the top ten obstacles in the popularity of cloud computing. The data security and storage issues is discussed in this article and it also analyzes the main reasons of data security issue, possible solutions of this issues and some future development of cloud computing are also discussed. [16] Explains the seven phase of data life cycle in cloud computing that also Need security to get user trust these phase include; generation, transfer, use, share, storage, archival and destruction. The aim of cloud...
computing is to provide better consumption of resources and reduce the work load from user end but it suffers with security threats [17]. The complexity of security in complete cloud computing environment is shown in figure 1. In figure 1 the lower layer indicates the deployment models of cloud computing namely private cloud, community cloud, public cloud and hybrid cloud. The layer just above the deployment model represents the services delivery model of cloud computing. These service delivery models exhibit the certain characteristics that are shown in the top layer. These fundamental elements need security with respect to the characteristics of selected deployment model. Some of fundamental security challenges are shown in the vertical layer given in figure 1.

In this research work we focused on the data security issue in cloud computing environment. Public cloud deployment model mostly suffers from the risk of data security. On the other hand, in SaaS delivery model client is dependent on service provider for proper security measures. The provider must implement some strict security measures to keep multiple users from seeing each other’s data and gain the trust of users. Recent reviews on security issues in cloud computing are presented in [21, 22, 23] but these reviews are limited and not focused on detail study of data security issue. Neither of them adopts a proper literature review process. In our study we focused in details study on data security issue by adopting a proper systematic literature review process.

3. METHODOLOGY

Empirical studies are now being undertaken more frequently, as a means of examining a broad range of phenomenon in computer field. A systematic literature review presented in [24] is followed in this research work to conduct the review. The review process is shown in figure 3. A systematic literature review endeavor to provide a comprehensive review of current literature relevant to a specified research questions.

Many researchers contribute their efforts in the field of software engineering/computer science by adopting [24] systematic literature review process such as in [25, 26] systematic literature review process is adopted for the review of aspect oriented implementation of software product lines components and software component reusability assessment approaches.
Phase 1: Plan Review
1- Specify research questions
2- Develop a review protocol
3- Validate review protocol

Phase 2: Conduct Review
4- Identify relevant research
5- Select primary studies
6- Assess study quality
7- Extract required data

Phase 3: Document Review
9- Write a review report
10- Validate report

Fig 3: Adapted review process from [24]

The review process has three phases that consist of ten sub activities. In first phase of review the following questions are posed:

**Question 1:** What approaches have been introduced to ensure data security in cloud computing?

**Question 2:** How the approaches have been validated?

The questions are formulated during the first sub activity of phase 1, a review protocol was developed. The review protocol includes the sources, time period under review and key words used. This protocol is reviewed and validated after making some changes by researchers.

The final review protocol is shown in table 1. The sources used for this review include science direct, IEEE explorer, Google scholar, Scopus, ACM portal digital library. Additionally we have looked at JCMS, IJSI journals. The research focuses on the year’s 2007 to 2020.

<table>
<thead>
<tr>
<th>Year</th>
<th>No. of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>0</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
</tr>
<tr>
<td>2009</td>
<td>1</td>
</tr>
<tr>
<td>2010</td>
<td>5</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
</tr>
<tr>
<td>2012</td>
<td>8</td>
</tr>
<tr>
<td>2013</td>
<td>9</td>
</tr>
<tr>
<td>2014</td>
<td>2</td>
</tr>
<tr>
<td>2015</td>
<td>7</td>
</tr>
<tr>
<td>2016</td>
<td>6</td>
</tr>
<tr>
<td>2017</td>
<td>4</td>
</tr>
<tr>
<td>2018</td>
<td>10</td>
</tr>
<tr>
<td>2019</td>
<td>9</td>
</tr>
<tr>
<td>2020</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
</tr>
</tbody>
</table>

Fig 4: Frequency of papers w.r.t to sources

In the second phase of review, the search is performed by using different queries related to data security in cloud computing environment. The initial collection of research papers was based on the key words in Table 1 in the papers keywords and abstract. The quality criteria set to assess the studies was to include papers in the review if it contains a model, an experiment, a framework, or a guideline. The required data was extracted from the papers to answer the questions posed above.

Another step in the search process was performed by searching the related work area of the selected papers to boost the review strength by confirming that no valuable reference is missed during the search process. The collected data was synthesized to exhibit complete results. Finally, in the third phase of the review process, the review report was written and validated.

4. RESULTS

The results of the review are presented in this section. A year wise result representation is presented in Table 2 and frequency of papers with respect to sources is shown in Fig 4. The results are characterized with respect to the questions posed earlier.

Table 2. Review protocol

<table>
<thead>
<tr>
<th>Year</th>
<th>sources</th>
<th>Key words</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-2020</td>
<td>IEEE Explore, science direct, Scopus, Google scholar, ACM portal digital library, IJERA, IJSI</td>
<td>Cloud computing, cloud computing security, data security/data concealment, cloud data security, cloud data storage</td>
</tr>
</tbody>
</table>

In the second phase of review, the search is performed by using different queries related to data security in cloud computing environment. The initial collection of research papers was based on the key words in Table 1 in the papers keywords and abstract. The quality criteria set to assess the studies was to include papers in the review if it contains a model, an experiment, a framework, or a guideline. The required data was extracted from the papers to answer the questions posed above.

Another step in the search process was performed by searching the related work area of the selected papers to boost the review strength by confirming that no valuable reference is missed during the search process. The collected data was synthesized to exhibit complete results. Finally, in the third phase of the review process, the review report was written and validated.

5. QUESTION-1: WHAT APPROACHES HAVE BEEN INTRODUCED TO ENSURE DATA SECURITY IN CLOUD COMPUTING?

The result of review (figure 5) show the proposed approaches for the data security in cloud computing. These results are categorized into: (1) Encryption, where the plain text is
5.1 Encryption

The results show that most common approach was encryption (45%) to assure the data security in cloud. In [27] a digital signature with RSA algorithm scheme is proposed to ensure the data security in cloud. In which software used to crunch the data and pick up the Unique Identification for Server in cloud server then it will check the message header of received message and this 64 bit plain text is divided into two halves by using the “black box” the right half have 2 bits whereas left half has 6 bits, then these 6 bits are fed into “superior function” block where these 6 bits are further separated in two halves where first two bits represent the rows and last four bits represent the column by identifying the rows and column the corresponding value can be selected. Then this function is applied to all 8 octets of the output of vigenere block the resultant of black box is again of 64 bits then these bits are further divided into 4 new octants similarly right 4 bits are unified to formulate right halves. Finally left and right halves are XOR-end to obtain left half of this arrangement. This process is repeated three times.

In [29] RSA algorithm used to encrypt the data and Bilinear Diffie-Hellman to insure the security while exchanging the keys. In proposed method a message header is added in front of each data packet for direct and safe communication between client and cloud without any third party server. When user sends the request to the cloud server for data storage then cloud server creates the user public key, private key and user identification in certain server. Two tasks performed at user end before sending the file to cloud, first add message header to the data and secondly encrypt data including message header by using secret key. When user request for data to the cloud server then it will check the message header of received data and pick up the Unique Identification for Server in cloud (SID) information. If SID information is found it will respond to the user request otherwise request will be discarded.

In [30], a technique is introduced to ensure the availability, Integrity and confidentiality of data in cloud by using Secure Socket Layer (SSL) 128 bit encryption that can also be raised to 256 bit encryption. The user who wishes to access the data from cloud is strictly required to provide valid user identity and password before access is given to the encrypted data. In [31], user send the data to the cloud then cloud service provider generate a key and encrypts the user data by using RSA algorithm and stored the data into its data center. When user request the data from cloud then cloud service provider verify the authenticity of the user and give the encrypted data to the user that can be decrypted by calculating the private key.

In [32], a three layered data security model is presented in which each layer performs different task to make the data secure in cloud. First layer is responsible for authentication. Second layer performs the duty of data encryption and third layer performs the functionality of data recovery. In [33], RC5 algorithm is implemented to secure the data in cloud. An encrypted data is transmitted even if the data is stolen there will be no corresponding key to decrypt the data. In [34] Role Base Encryption (RBE) technique is proposed to secure the data in cloud and role base access control (RBAC) cloud architecture was also proposed which allows organizations to store data securely in public cloud, while maintaining the secret information of organization’s structure in private cloud.

In [28] play fair and vigenere cipher techniques were merged with structural aspects of Simplified Data Encryption Standard (SDES) and Data Encryption Standard (DES). In which 64 bit block size of plain text is taken which is fixed and this 64 bit plain text is divided into two halves by using the “black box” the right half have 2 bits whereas left half has 6 bits, then these 6 bits are fed into “superior function” block where these 6 bits are further separated in two halves where first two bits represent the rows and last four bits represent the column by identifying the rows and column the corresponding value can be selected. Then this function is applied to all 8 octets of the output of vigenere block the resultant of black box is again of 64 bits then these bits are further divided into 4 new octants similarly right 4 bits are unified to formulate right halves. Finally left and right halves are XOR-end to obtain left half of this arrangement. This process is repeated three times.

In the studies have outlined some guidelines to ensure we can directly compare with encrypted token; (3) Guidelines. Building Some of the studies have outlined some guidelines to ensure we do not need to decrypt the key for data checking instead we can directly compare with encrypted token; (3) Guidelines.

Table 3 category wise results of question1

<table>
<thead>
<tr>
<th>Question</th>
<th>category</th>
<th>No. of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>What approaches have been introduced to ensure data security in cloud computing?</td>
<td>Encryption</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Homomorphic token</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Sabol sequence</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Guideline</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Harmonizing scheme</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Data concealment component</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Framework</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Stripping algorithm</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>31</td>
</tr>
</tbody>
</table>

Fig 5: Proposed approaches to ensure data security

Produce the digital signature. Digital signature will be decrypted into message digest by the software with own private key and public key of sender.
In [35], four authorities are defined i.e., data owner, data consumer, cloud server and N attribute authorities where attribute authorities sets was divided into N disjoint sets with respect to the category. The data owner gets the public key from any one of the authority and encrypt the data before sending it to the cloud server. When data is requested the authorities will create private key and send it to the data consumer and consumer will be able to download the file only if he get verified by cloud server. In [36], two types of secure cloud computing are proposed one require trusted third party and other does not. These types use Elliptic Curve Diffie-Hellman (ECDH) and symmetric bivariate polynomial based secret sharing to ensure the data security in cloud environment.

In [37], location based encryption technique by using user location and geographical position was introduced. In which a geo encryption algorithm was implemented on the cloud and user computer and the data was labeled with the company name or person who work in the company. When the data is required then in the cloud similar label will be searched and retrieved and the information corresponding to the label will be retrieved. In [38], a technique is proposed by using digital signature and Diffie Hellman key exchange in combination with Advanced Encryption Standard encryption algorithm to protect the confidentiality of data stored in cloud. This scheme is referred as three way mechanism because it provides authentication, data security and verification at the same time.

5.2 Guidelines
The result of our review shows that 21% of studies use guidelines to ensure the security of data in cloud. In [39], guidelines are provided for data security in cloud by introducing new cloud system architecture approach which has three features i.e., separation of software service providers and infrastructure service providers, hiding information about owner of data and data obfuscation. In [40], agents method is introduce to ensure the data security in cloud architecture. In which three agents namely file agent, authentication agent and key managing agent was used for data security.

In [41], guidelines about six key data technologies are provided which are: data privacy protection, proof of existence and usability of data, trusted access control, retrieve and process of cipher text, cloud resource access control and trusted cloud computing. In [42], guidelines are provided by giving the meta-analysis of four different encryption algorithms that are also helpful to selecting the best algorithms according to need.

5.3 Framework
The framework approach represents 14% of the results. In [43], a framework is provided; known as Trust Cloud, in which data centric and detective approach is propose to increase the security of data with the objectives to encourage the adoption of file-centric and data-centric logging mechanism to increase the security and confidentiality of data in cloud computing. In [44], a framework is provided by building a multi-tenant system. In which developed solution is divided into three layers i.e. presentation layer, business logic layer and data access layer. These layers provide very high security to user data.

In [45], a framework is provided that consists of protocol named Sec Cloud, which is a first protocol spanning secure storage and secure computation in cloud environment by designated verifier signature, batch authentication and

Probabilistic sampling procedures. In [46], proposed framework is consist of three steps, in first step precaution is made against semi-honest cloud service provider by indexing data and its metadata to ensure complete data privacy. In second step multi user private keyword searchable encryption is performed on encrypted data to keep searches and resulting files secrecy from cloud service provider. Final step make the use of policy in order to support data sharing between users by using metadata and encryption scheme.

5.4 Homomorphic Token
The homomorphic token scheme represents the 7% of the results. In [47], homomorphic token scheme is introduced to ensure the data security. The proposed scheme utilizes homomorphic token with distributed verification of erasure-coded data. It supports secure and efficient dynamic operation on data block including data delete, update and append. A model proposed in [48] by utilizing homomorphic token scheme with token pre-computation algorithm to achieves the integration of storage correctness insurance and identification of misbehaving server(s).

5.5 Stripping algorithm, data concealment component, harmonizing and token scheme
Stripping algorithm, data concealment component, and harmonizing and token scheme each represent 3% of the results. In [49], stripping algorithm is used to secure the picture data in cloud, the approach is consist of three modules which are image analysis, data separation and data distribution. [50] Proposed a design of data concealment component that composed of three sub components: the prediction component, data generator and data marking to secure the data in cloud. The Evaluation of this component shows the successful conceal data of legitimate users and protect them against potential attacks.

A privacy preserving repository presented in [51], this repository was basically concentrated on the harmonizing operations to achieve data confidentiality while still keeping the harmonizing relations intact in the cloud. This proposed scheme make data owner enables to assign most of computation intensive tasks to cloud servers without disclosing data contents. [52] Proposed an effective and flexible distribution verification protocol to address data security in cloud computing. This protocol utilizes token pre-computation using sobol sequence to verify the integrity of erasure coded data instead of pseudorandom data. The proposed model consist of three phases that are: file distribution, token pre-computation and challenge response protocol.

6. QUESTION-2: HOW THE APPROACHES HAVE BEEN VALIDATED?
The results related to the second question are presented here. Figure 6 shows the result of review regarding the procedures adopted for validation. The categories are: (1) Experiment, where an experiment is carried out to validate the results; (2) Comparative analysis, where the results of proposed scheme is compared to other schemes to validate the results; (3) Test bed is used to validate the proposed approach; (4) Statistical analysis, where the results are analyzed by using some statistical technique; (5) Meta-analysis is used to validate the
Results: (6) Performance analysis, where the performance of proposed approach is analyzed by different methods; (7) some of the proposed approaches have not performed any validation. The category wise detail is presented in table IV and fig 6 shows the type of validation in percentage. Let us explain the term validation. It refers to any kind of empirical method used as a proof, apart from the demonstration/application of the proposed approach.

<table>
<thead>
<tr>
<th>Question category</th>
<th>No. of papers</th>
</tr>
</thead>
<tbody>
<tr>
<td>How the approaches have been validated?</td>
<td></td>
</tr>
<tr>
<td>Experiment</td>
<td>10</td>
</tr>
<tr>
<td>No Validation</td>
<td>13</td>
</tr>
<tr>
<td>Comparative Analysis</td>
<td>3</td>
</tr>
<tr>
<td>Meta-Analysis</td>
<td>1</td>
</tr>
<tr>
<td>Test Bed</td>
<td>1</td>
</tr>
<tr>
<td>Statistical Analysis</td>
<td>1</td>
</tr>
<tr>
<td>Performance Analysis</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 4 categories wise results of question 2

Fig 6: Type of validation

The results of the question regarding validation of proposed approaches show that 47% of the selected papers proposed approach to secure data in the cloud environment but provide no validation of the proposed approach.

6.1 Experimental Approach
Experiments are used to validate the proposed approach in 32% of the selected papers. In [30], experiment was performed to test the validity of proposed model by using cloud simulator named Hadoop. It shows the status of security after implementing three security parameters which are; Message Authentication Code, classification of data, index and encryption technique. In [33] Aneka 2.0 software is used in cloud environment to validate the results obtained by the implementation of RC5 algorithm and then compare these results with Amazon S3 service. Aneka allows building and managing an interconnected network by using Microsoft.NET frameworks on these networks.

In [34] proposed architecture is implemented in Java and results show that cipher text size is linearly proportional to the size of the plaintext and the efficiency of encryption and decryption is very good. Results also show that the size of the decryption key is 48 bytes which is convenient for the users. In [39] cloud service is implemented using C# Microsoft .NET framework for collaborative online documentation. The experimental results shows that service response time increases linearly as the size of the input text increases and data obtuse-cation and de-obfuscation do not cause much overhead, hence proposed approach showed realistic performance. In [50] PHP language was used for the experiment in which performance test is conducted for three phases that are data generation, data marking and data extraction. During the performance test impact of component on data generation was also observed.

6.2 Comparative Analysis
Comparative analysis as the form of validation is employed in 10% of the selected studies in which results of proposed scheme is compared to other schemes to validate the results. In [53] comparative analysis is conducted to validate the results by considering following variables granularity, key management, meta data management, level of concealment, degree of distribution and level of implementation. In [36] comparative analysis is made between data privacy by Authentication and Secret Sharing (PASS) and proposed technique that used trusted third party and non-trusted third party. In [28], the proposed encryption technique is compared with DES, SDES, Play fair and Vigenere encryption technique to validate the proposed approach results.

6.3 Performance Analysis
Performance analysis is used to validate the proposed approach in 7% of selected papers. In [47], performance analysis is performed in terms of security and efficiency to show that the results are validated and proposed scheme is highly efficient and flexible against Byzantine failure and malicious data modification attacks.

6.4 Statistical Analysis
Statistical analysis, meta-analysis and test bed as the form of validation are employed in 3% of the selected studies. In [32] NIST statistical test are used to validate the results by selecting eight modern encryption algorithms. In [42], meta-analysis of four different security algorithms which are; AES, RSA, Blow fish and DES are presented in term of platform, key size, key used, scalability, initial vector size, security, data encryption capacity, authentication type, memory usage and execution time to validate the results. In [45], test bed is developed and tested for the validation of results.

7. CONCLUSIONS AND FUTURE DIRECTIONS
There are many benefits of using cloud computing such as cost efficiency, quick deployment, improved accessibility etc. However, there are yet many practical problems which have to be solved. The data confidentiality is one of them. Many...
Researchers contributed their efforts to minimize the data security issue in this domain with different solutions that described in this work. A literature review of the works in the area of cloud computing data security is conducted and the results of review are presented in this paper. The results show that the majority of approaches are based on encryption (45%) out of which 71% encryption techniques results are validated. 67% of encryption techniques used experimentation to validate the results. These results point towards the fact that most of researchers show their interest in encryption technique to enhance the security of data in cloud computing environment. The results also reveals the fact of lack of validation in proposed approaches as 42% of the studies provide no validation of the results out of which 67% are guidelines. Only few studies have used statistical analysis for validation. This area (validation) needs the attention of the research community to gain the trust and confidence of cloud computing users.

Although our review has explored the field, further studies are needed to confirm the obtained results. Future work includes the extension of this review by including more sources (conferences, journals and workshops) and questions. A future plan is to explore the other security issues in the cloud computing environment and we are also aiming to design a security model using some encryption techniques for data concealment in cloud computing.

8 REFERENCES


