

Comprehensive Study on Data Security in Cloud Data Store

Hisham A. Shehadeh, Qusai Y. Obeidat, and Derar Darwish

Department of CS, CIT Faculty, Just University, Jordan
e-mail: hashehadeh12@cit.just.edu.jo

Department of CS, CIT Faculty, Just University, Jordan
e-mail: qusai.gts@gmail.com

Department of CS, CIT Faculty, Just University, Jordan
e-mail: derardrweesh@yahoo.com

Abstract

There are many hot topics related to cloud computing paradigm, and one of these important fields is data security in cloud data storage that aims to make users data more secure and for ensure privacy for users. In this paper we intend to introduced a comprehensive study on security of cloud data store, since there have been little works deals with. The introduced papers here addressed this problem from different views, some deals with encryption algorithm while other researchers made comparisons between different techniques. On the other hand, some papers proposed new technique or made some enhancements on existing ones either depends on statistical or un-statistical encryption algorithms. Till now there are no optimal suggested solutions that solve the security on cloud data store.

Keywords: *cloud Computing, Encryption Algorithm, Cloud Storage, Data-Base, and Security.*

1 Introduction

Cloud computing is a computing technology that depends on the internet for providing the users with re-sources such as infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), and software-as-a-service (SaaS) [3].

Cloud computing has many advantages, some of them are providing the customers with computers in a lower cost higher performance, lower cost of software's, and infinite storage capacity etc, some of the disadvantages of the cloud computing are the need for constant internet connection, the need for high speed connection, customers data stored in a cloud data stores may be not secured [1].

Large number of users put their data in a cloud data stores, as the increasing amount of data stored in these data stores, this leads for locking in the most challenges faced the propagation of cloud computing, which is the security problem, such as the correctness, integrity of a data stored in the cloud, which considered as a main is-sue concerned by the user [2].

In this work we're concerned in the security problem that faced the propagation of the cloud computing, by providing some of the algorithms used in order to make the user's data that stored in the cloud data stores, to be secured. A comparison between these algorithms made in terms of the degree of security provided by each algorithm for the users' data. We organized this paper as follows Section II presents related works. Section III presents the comparison on introduced papers. Section IV presents conclusion. Section V presents open problem.

2 Related work

Although there have been many works on data security in cloud data store through the last decade, these works used a different type of encryption algorithm. We mentioned some of them.

Jay Singh, et al. [3] proposed secure cloud computing data store using Rc5 encryption algorithm to provide security and privacy for user data on cloud storage.

In the first section of their paper they proposed a cloud computing services such as Infrastructure as a Service (IaaS): it is providing computers, servers, and file storage for users, Platform as a Service (PaaS): it is including operation system, web server application and data-base, Software as a Service (SaaS): it is providing any type of application software. By these services users can use any software and store their data and file on cloud data storage. Authors used an Rc5 encryption algorithm with cloud computing to make more privacy and security on user data from hackers and spicily from cloud storage providers because they can access user's data directly.

They tested their technique using ANEKA 2.0 cloud environment with Microsoft visual studio to implement Rc5 algorithm on cloud data storage. The rustles showed that the system worked well with high performance and high security because only the user has and knows the decryption key.

Rewagad, et al. [4] proposed data integrity and confidentiality in cloud data store for users and corporations by using two types of security levels; login key encryption using Hellman key encryption algorithm and data encryption using Advanced Encryption Standard encryption algorithm (AES).

Cloud computing provide a set of service for users and corporations such as Infrastructure-as-a-Service (IaaS), Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS). These services are available via network so, users' data may possible to leakage and steal by hackers. In the first section of their paper they proposed a set of attacks on cloud data store and a solution for prevention the break of users' privacy. The first attack is Tampering: attacker can alter the data in cloud data store, the second attack is Worms and Viruses: attacker can corrupted data files in data store and decries the performance of hardware; the third attack is Elevation of Privileges: attacker can unauthorized access of user data and modified on it.

They solving these problems by using two type of servers, one to encrypt/decrypt data files that uploaded by user using AES encryption algorithm after check login authentication by Hellman key encryption, the another server is used to store encrypted user files. The result showed that the system can be applied without need high hardware requirement such as 128 megabyte RAM, 1.2GHz processor.

P.Varalakshmi, et al. [5] proposed secure cloud computing data store using RSA data encryption algorithm to provide security and privacy for user data on cloud storage.

Cloud remote server may be under attack by hackers this breakable the privacy of user or organization data by steels it. Because of this, authors used a set of layer of protection organization or users' privacy. The first layer is broker layer (third party) in which is used as the interface between user and storage layer, the security is checking and the encryption and decryption is done in this layer. The second layer is storage server to store encrypted user data that come from broker. The third layer is user layer: user is store or retrieves data on cloud server; broker checked the users' authentication after that retrieves data by decrypt it to users or store data by encrypt it in storage server.

They tested their system on 50 files using RSA encryption algorithm and Xen operation system in which allow many VMs (storage server). The result showed that the system given better encryption and decryption time using RSA algorithm.

Mrs,G.Nalinipriya et al. [6] proposed secure cloud computing data store using AES-NI encryption algorithm with hash function algorithm MD5 to provide security and privacy for user data on cloud storage.

In first section of their paper they proposed a type of cloud computing such as public clouds, hybrid cloud, and private clouds. These types of clouds provide a set of services such as Infrastructure as a Service (IaaS) in which provides computers, servers, and file storage for users, Platform as a Service (PaaS) in which includes operation system, web server application and database, Software as a Service (SaaS) in which provides any type of application software. These services allow users to upload their data on cloud storage servers; these servers may be under attack by hackers this breakable the privacy of user or organization data by steals it.

Authors used AES-NI encryption algorithm with MD5 hash algorithm to trust security on user or organization data. Hash function is used to encrypt user key that is used in AES-NI encryption algorithm. When user is up-loading his/her data on server the system is required from user a hash number to access AES-NI encryption algorithm, after that the system is encrypt data and stored it in storage server with hash number. When users wants retrieve their data the system is request from user to input hash number, if the hash number that is inputted by user is the same number that stored with encrypted data the system is allow AES-NI algorithm to decrypt data else the access is not allow. They tested their system using laptop with Core i7 CPU and the result showed that the system is work speed.

Parsi Kalpana, et al. [7] provided a method to solve the security challenges that faces the propagation of the cloud technology as the large amount of data stored in the cloud stores made the security issues as the main problem that must be solved, since the main concern of users and companies is to keep their data secured from hacking challenges, they introduced in the first section some of the security issues concerned by the user, such as privacy and confidentiality which means that the user required that accessing of data should be determined only for the authorized person.

Data integrity is considered as another security issue that concerned by the user such that the cloud providers must provide security methods that guarantee the integrity of users data, another security issue is the data location since the users don't know the place of their data in the cloud ,and what servers their data resides in, unless the user agreement with the cloud provider favorite place for his data through an determined a on the other hand, the movement of user's data from one location to another, is another risk for the user as his data firstly stored in a particular place, and then it moves from one place to another. Data availability is another issue concerned by the user, such as the data of the user is often stored in different places or clouds but the permanent availability of user's data is very difficult.

Storage, backup and recovery is also very important issue for the user such as the cloud provider often saves user's data in different servers with multi copies, and options for backup is provided specially for businesses, who worked on the cloud applications ,so that they can returned back to the previous state if a failure of hardware occurred. In their proposed approach, they used the RSA algorithm for improving the security of user's data.

RSA algorithm works in the following steps:

1. Key generation: in this process two keys are generated the first one is the public key, which is used by cloud provider to encrypt data, and the second one is the private key that is generated only for the user ,who can decrypt the data by using it.
2. Encryption: in this step the original text of the user is converted into a cipher text by using the public key.
3. Decryption: in this step the cipher text is converted to the original text using the private key of the user.

Amandeep Kaur¹, et al. [8] provided a new way to protect the user's data to be secured in the cloud data store using RSA & Twofish algorithms. In the first section of their paper they described the cloud computing technology, which provides the users with services and resources over the internet, such as the user stores his data in the cloud data store instead of storing the data in the hard drive of his personal computer. Another point that the cloud computing model provides, a way for accessing the information and resources from any place, where the connection of the internet is available. Cloud computing provides a way for sharing resources of the cloud such as data storage space, processing power and applications.

In addition, they mentioned some of the cloud computing benefits such as the cloud technology reduces the costs, paid by the companies and saving them for increasing the computing capabilities, and also reducing supporting costs of the company system.

Another benefit is the flexibility such as the cloud computing allows the companies to use the resources as much as they need. Reliability is another benefit when the services of the cloud existing on many servers, this gives the company a reliability for its work if a disaster occurred. Maintenance is another benefit, as it is the responsibility of the cloud provider, so the cost of maintenance is reduced. Mobile Accessible is very important benefit; as it increases the productivity of workers.

In the second section of their paper they mentioned the challenges that face the propagation of the cloud technology. Some of them are security and privacy which considered being the most critical issue that limits the propagation of the

cloud. Continuously Evolving is another challenge, such as the continuously increasing of the user's requirements forcing the public cloud to evolve continuously.

User's data security is very important issue, so in order to keep the user's data to be secured the authors provided a model for improving the security on data storage. The model is simulated by using the visual studio environment using the Azure Cloud. The proposed model follows the following steps in order to increase the user's data security:

1. RSA algorithm is implemented for Encryption and Decryption of data.
2. The TWOFISH algorithm is also implemented for Encryption and Decryption of data.
3. In order to send the private key to the user's id mail method is used.
4. Increasing the security of data by adding signature.

By using this proposed approach, the user can upload the data file, and it will be encrypted when it is stored in azure cloud, by using the two algorithms RSA and TWOFISH .in addition, the data file will be locked using a signature.

In this approach when the user upload the data file, RSA and TWOFISH algorithms is used to encrypt the data file, and generating a key and a signature for blocking the data file, and when downloading the data file, the opposite way will be as the RSA and TWOFISH algorithms will be used for decryption, and the signature will be used to unlock the data file, so that only the user who own the valid key and signature can download a data file.

The results of their proposed approach showed that the security level was higher than the Previous Approach.

Debajyoti Mukhopadhyay, et al. [9] provided a framework for keeping files to be secured from possible attacks, by using file encryption operation using the AES algorithm. In the first section of their paper, they described the cloud computing technology which provides many services to users. It provides a large storage space and a high computation speed for the user. In the other hand, they described the most important obstacle that faces the propagation of the cloud computing technology, which is the security issue such as the integrity and availability of user's data. The proposed technique relies on encrypting the users' files to be uploaded in the cloud, and ensuring the integrity by allowing access to these files by only the Authorized user. The second section of their paper mentioned the services provided by the cloud computing which they are:

1. Storage as a Service: this service enable the user to have a very large space for storing data, and the cloud systems must provide techniques to the users' data to be secured.

2. Software as a Service: this service provides the user the ability to access software's that resides in the cloud, from any computer that support a browser without the need to install software's in the personal computer.
3. Platform as a Service: this service provides the user with the ability to create software's by using the cloud libraries and tools, so that costs related to underling software and hardware will be reduced.
4. Infrastructure as a Service: this service provides the equipment's needed to serve the user such as hardware and network components.

Many security issues are mentioned in this paper such as data breaching as it considered as the largest security problem, as the hacker can reach the client application, and then to the client privacy information .another security issue is Inefficient APIs which is considered as an easy goals for attackers. Another issue is Denial of Service where the user provided partially accessing to data or it can't access to his data. Connection eavesdropping, loosing of user's data, compatibility that it issues faces the cloud computing technology.

The proposed approach securing files by encrypting files to be uploaded by using a password generated using the AES algorithm, and the encrypted files can be also downloaded by the user.

AES algorithm is used because of the following reasons:

1. AES algorithm can't be attacked because the key used for AES algorithm consists of 128 or 192 or 256 bit.
2. When we comparing AES algorithm to the RSA algorithm AES algorithm is considered to be faster than RSA, so that we use it to secure the user's data instead of RSA.

The following steps are considered when uploading files :

1. The system checks the validity of the user name and password for the user if they are true, then a connection with the cloud will be created else an authentication error will be established.
2. In this step when the user is authenticated and a connection is created with the cloud, then the user must select any file resides in his machine.
3. A long password is recommended to be selected to be used for generating the key needed for encrypting files.
4. This step generate the key from the password where the key is used for encrypting data, and AES algorithm is considered as symmetric key algorithm as the same key is used for the encryption and decryption process.
5. In this step, the AES algorithm is applied for the encryption process, so that the plan text is transformed to the cipher text using this algorithm.

It must be noticed that the user's data is doubly protected as the username and password are existed to ensure the validity of user, and the user's data also is encrypted.

The following steps are considered when downloading files:

1. As the first step in uploading a file the validity of username and password is checked for the user.
2. In this step all the files that has been uploaded by the user is appeared to the user and the user must select one of them.
3. In this step the user must enter the same password used in the encryption process.
4. This step validate the password entered by the user, so that the cipher text will be decrypted, only if the user used the same password used in the encryption process.
5. In this step we use AES algorithm in order to decrypt the cipher text by using the generated key.
6. The plan text is stored in the user machine memory.
7. Finally the user may be asked if he wants to delete the encrypted file, so if he wants to delete it then the encrypted file must be deleted. When the user doesn't need to download any other files, then he must logging out and breaks the connection with the cloud.

S.Ezhil Arasu, et al. [10] a new strategy to ensure the correctness of data in the cloud data store. In the first section of their paper, they mentioned the benefits of the cloud computing technology, such as maintenance and updating of applications that is done by the cloud provider. In addition, they mentioned that in order to ensure the correctness of data TPA (third party auditor) is used. Their new approach is focusing on using the HMAC (Hash Message Authentication Code) it is a cryptographic function that makes a concatenation between the message and a secrete key, in order to generate an authentication code which is created by using a hash algorithm, as SHA algorithm .the integrity of the user's data is done by using the authentication code. the hash algorithm like SHA create a separate key and passing the key with the original message of the user to generate the authentication code, and a comparison is made between this authentication code of the user and the other one, generated by the auditor, so that the auditors use this algorithm which is called HMAC, in order to ensure the correctness of the user's data.

Arjun Kumar, et al. [11] proposed a Secure Storage and Access of Data in Cloud Computing. The aim of this work is to allow user to store and to access the data from the cloud storage securely. They propose a method that build a trusted environments, the method are encrypt data before send it to storage and decrypt data after receiving it from storage in client side using secret key. In their model they proposed that the cloud storage server have two parts: private data section and Shared data section. They use elliptic curve cryptography encryption (ECC) technique to encrypt data in two sections. The data of private section is encrypted using ECC private key and the data of public section is encrypted using ECC public key. The proposed method have four phases authentication (user must enter

username and password), operation (ECC generate secret key using pin number), Encryption (ECC encrypt data using secret key that need to store on cloud) and Decrypt (decrypt data that downloads from cloud to original file). This method assumes the storing and accessing data is much secure.

Chao YANG, et al/ [12] proposed A Novel Triple Encryption Scheme for Hadoop-based Cloud Data Security. The goal of this research paper is to ensure data security in cloud data storage. They combine HDFS encryption using DEA (Data Encryption Algorithm) and the data key encryption with RSA, and then they encrypt the user RSA private key using IDEA (International Data Encryption Algorithm). They use hybrid encryption method to encrypt HDFS files. With hybrid encryption method the files are encrypted using DES algorithm and get the data key and then data key is encrypted using RSA algorithm to get a private key.

The user use private key to decrypt the data key. The hybrid encryption method is Improving the security but the private key of the user may be stolen by attackers. In this paper authors also encrypt data private key using IDEA. They evaluate the performance overhead by the triple encryption scheme into HDFS they use three cluster scenarios. The first is Pseudo-distributed model with replica=1, Fully-distributed model with replica=1 and Fully-distributed model with replica=3. They conduct 12 different files for each scenario and each scenario take 10 times. Then they calculate average rate of Input/output of default HDFS and triple HDFS scheme. They are analyzing file writing (uploading files from client to HDFS) and file reading (downloading files from HDFS to client). They show that the uploading process and the downloading process rate of the triple HDFS encryption scheme is slower than default HDFS because operation of triple encryption scheme.

Arjun Kumar and HoonJae Lee [13] proposed a method that provides a secure platform in Cloud Computing to builds a trusted computing environment. The method they proposed allows users safely and efficiently to store data. The security issues and handling big data problem is solved using encryption and compression technique when that uploading to the cloud storage. In this model the user interact with main server to retrieve data and upload data, when he upload data to the main server the backup process is take place.

They use three backup servers to recover data if the disaster is happen. If one of server path fails it will use the alternate path to process data. Encryption and compression process of files takes place during backup, decryption and decompression during recovery. The important security services in this method are authentication, encryption, decryption and compression and decompression. Each user has a secret key that is generated by the main server. User login to the main server using his email ID and password, if this step is done successfully then the main server is must authorize person login using email as ID and secret key as a password. This two times authentication will improve the security level.

Shuai Han, et al. [14] proposed a novel third party auditor scheme. This scheme aims to ensure data stored security in cloud and access control safely. In this scheme they use RSA and Bilinear Diffie-Hellman. They encrypt data between servers using RSA algorithm and the Bilinear Diffie-Hellman will insure security when keys are exchanging. Each user has pair of key that is used for cloud access control. The user adds a message header for each data file that is encrypting using RSA.

They partition their scheme into two parts, the first part is users and cloud storages servers and the second one are cloud storages servers and organization server that are trustful servers. In the first part if the user wants to upload data file he sends a request to any one of cloud storage servers then the server creates a pair of key that are user public key (UPK) and user secret key (USK). The server then encrypt data file before uploading it into cloud storage server. The interoperation between cloud storage servers and cloud service providers is done in a trust state using unique identification for server (SID) in cloud for each user. In this scheme the performance metrics that are analysis are Access control that the user without privilege can't access to other function that he must have a privilege to access it. The other performance metrics are Authentication data trustful and authenticate efficiently.

3 Comparison

In this section, we make a comparison between all papers studied here in the review of literature that we have already made in the previous section. To make this study, we take in consideration the following factors: (1) The encryption type that is used data, file or key (2) Proposed Method that had been used in each paper (usually referred to the method name or technique used). (3) The test environment if it mention in paper. In table1 we represent some of comparison critics about proposed methodology. Comparison factors are index type, model, statistical, un-statistical, proposed methodology, statistical computations.

Table 1: Comparison Methodology

<i>Paper #</i>	<i>Encryption Type</i>	<i>Proposed Methodology</i>	<i>Test Environment</i>
[3]	Data encryption	Used Rc5 encryption algorithm with cloud computing data storage to make more privacy and security on user data.	ANEKA 2.0 cloud environment

[4]	Data encryption	Using two types of security levels; login key encryption using Hellman key encryption algorithm and data encryption using Advanced Encryption Standard encryption algorithm (AES).	Laptop with 128 megabyte RAM, 1.2GHz processor
[5]	Data encryption	Used a set of security layer: broker layer to encrypt/decrypt data, storage layer to store encrypted data.	Xen operation system: VMs represent storage servers, host is broker.
[6]	Data encryption	secure cloud computing data store using AES-NI encryption algorithm with hash function algorithm MD5.	Laptop with Core i7 INTEL CPU
[7]	Data encryption	using the RSA algorithm for improving the security of user's data in the cloud computing storage	-
[8]	File encryption	Using RSA and TWOFISH algorithms in order to encrypt the data file in addition the security of the data file is increased by locking the data file by a signature.	Visual studio environment using the Azure Cloud.
[9]	File encryption	Using AES algorithm to encrypt the user's files and user's data is doubly protected since username and password are checked for the user when creating a connection with the cloud for uploading or downloading process in addition to user's data encryption.	-
[10]	Message authentication code	Using HMAC algorithm in order to generate the authentication code that is made by making concatenation between the message of the user and a secret key generated by a hash algorithm like SHA algorithm and making comparison between the user authentication code with the TPA authentication code to ensure the correctness of data.	-
[11]	Data encryption	Using elliptic curve cryptography encryption	-

[12]	Data and key encryption	They combine HDFS encryption using DEA and the data key encryption with RSA, and then they encrypt the user RSA private key using IDEA	Using three cluster scenarios. They conduct 12 different files for each scenario and each scenario take 10 times.
[13]	File encryption and compression	using encryption and compression technique when that uploading to the cloud storage	-
[14]	Data encryption	Using RSA and Bilinear Diffie-Hellman. They encrypt data between servers using RSA algorithm and the Bilinear Diffie-Hellman will insure security when keys are exchanging.	-

4. Conclusion

As we make a comprehensive study in this paper, we noted that have been little works deals with the security issues that limits the propagation of the cloud computing technology. Till now there is no optimal suggested solution that solves the security problem of the cloud data store. For this reason and others we intended to make our work as a survey paper to make it easier for anyone interested in working on this field gathering as many information and references as we could introduced here.

5. Open Problem

In this article comprehensive study on data security in cloud data store has been presented, since there have been little works deals with. We make a comparison between a few selected papers. A comparison between these papers made in terms of the degree of security provided by each algorithm for the users' data. These papers addressed this problem from different views, some deals with encryption algorithm (statistical or un-statistical encryption algorithms) while other researchers made comparisons between different techniques. On the other hand, some papers proposed new technique or made some enhancements on existing ones. Till now there are no optimal suggested solutions that solve the security on cloud data store.

References

- [1] Shivaji p . Mirashe,Dr. N.V.Kalyankar," Cloud Computing", JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.
- [2] Miss. M.Sowparnika1, Prof. R. Dheenadayalu2,"Improving data integrity on cloud storage services", International Journal of Engineering Science Invention, ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726, www.ijesi.org Volume 2 Issue 2 | February. 2013 | PP.49-55.
- [3] JJay Singh, Brajesh Kumar, Asha Khatri," Improving Stored Data Security In Cloud UsingRc5 Algorithm ", Conference on Engineering (NUiCONE), 2012 Nirma University International.
- [4] Mr. Prashant Rewagad, Ms.Yogita Pawar," Use of Digital Signature with Diffie Hellman Key Exchange and AES Encryption Algorithm to Enhance Data Security in Cloud Computing ", Conference on International Communication Systems and Network Technologies. 2013.
- [5] P.Varalakshmi, HamsavardhiniDeventhiran" Integrity Checking for Cloud Environment Using Encryption Algorithm ", Conference on Recent Trends In Information Technology (ICRTIT), 2012International 19-21 April 2012.
- [6] Mrs,G.Nalinipriya ME., (Phd), Mr.R.Aswin Kumar," Extensive Medical Data Storage WithProminent Symmetric Algorithms On Cloud - A Protected Framework", 2013 International Conference on Smart Structures & Systems (JCSSS-20 13), March 28 - 29,2013, Chennai, INDIA.
- [7] Parsi Kalpana ,et al," Data Security in Cloud Computing using RSA Algorithm", International Journal of Research in Computer and Communication technology, IJRCCT, ISSN 2278-5841, Vol 1, Issue 4, September 2012.
- [8] Amandeep Kaur1, Sarpreet Singh2," Improved Storage Security Scheme usingRSA&Twofish Algorithm at Window Azure Cloud", International Journal of Computer Trends and Technology (IJCTT) – volume 4 Issue 7–July 2013.
- [9] Debajyoti Mukhopadhyay,et al," enhanced security for cloud storage using file encryption",Cornell University, arXiv preprint arXiv:1303.7075 (2013).
- [10] S.Ezhil Arasu,B.Gowri,S.Ananthi," Privacy-Preserving Public Auditing In Cloud Using HMAC Algorithm", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-2, Issue-1, March 2013.
- [11] Arjun Kumar, Byung Gook Lee, HoonJae Lee, and Anu Kumari, “Secure Storage and Access of Data in Cloud Computing” , ICT Convergence (ICTC), 2012 International Conference on. IEEE, 2012.
- [12] Chao YANG, Weiwei LIN and Mingqi LIU, “A Novel Triple Encryption Scheme for Hadoop-based Cloud Data Security”, 2013 IEEE Fourth International Conference on Emerging Intelligent Data and Web Technologies.
- [13] Kumar and HoonJae Lee, “Efficient and Secure Cloud Storage for Handling Big Data”, Information Science and Service Science and Data

- Mining (ISSDM), 2012 6th International Conference on New Trends in. IEEE, 2012
- [14] Shuai Han, Jianchuan Xing, "Ensuring data storage security through a novel third party auditor scheme in cloud computing." Cloud Computing and Intelligence Systems (CCIS), 2011 IEEE International Conference on. IEEE, 2011