Text File Privacy on the Cloud Based on Diagonal Fragmentation and Encryption

Tawfiq Barhoom and Mahmoud Y. Abu Shawish

https://doi.org/10.33976/JERT.8.1/2021/3

Abstract— Despite the growing reliance on cloud services and software, privacy is somewhat difficult. We store our data on remote servers in cloud environments that are untrusted. If we do not handle the stored data well, data privacy can be violated with no awareness on our part. Although it requires expensive computation, encrypting the data before sending it appears to be a solution to this problem. So far, all known solutions to protect textual files using encryption algorithms fell short of privacy expectations. Thus is because encrypting cannot stand by itself. The encrypted data on the cloud server becomes full file in the hand causing the privacy of this data to be intrusion-prone, thus allowing intruders to access the file data once they can decrypt it.

This study aimed to develop an effective cloud confidentiality model based on combining fragmentation and encryption of text files to compensate for reported deficiency in encryption methods. The fragmentation method used the strategy of dividing text files into two triangles through the axis. Whereas the encryption method used the Blowfish algorithm.

The research concluded that high confidentiality is achieved by building a multi-layer model: encryption, chunk, and fragmentation of every chunk to prevent intruders from reaching the data even if they were able to decrypt the file. Using the privacy accuracy equation (developed for the purpose in this research), the model achieved accuracy levels of 96% and 90% when using 100 and 200 words in each chunk on small, medium, and large files respectively.

Index Terms— Cloud Computing; symmetric encryption; Cloud confidential; privacy; Fragmentation.

I INTRODUCTION

Organizations use clouds for various purposes such as organizational communication (Email), Services, and archiving correspondence to maintain mobility and efficiency with regard to organizational resources and cost across multiple platforms. Nevertheless, cloud is risky. This is because it is vulnerable to unauthorized access, which is potentially harmful to organizations in ways that cause serious problems. Therefore, it is highly critical to protect the privacy of cloud data through using the proposed combined model herein referred to as Protected Textual Cloud Model.

The issue of protecting the privacy of files on the cloud environment has become a problem that affects all companies and institutions. The mounting need to use clouds resulted from the cheapness of its use compared to building an integrated server room associated with the human resources necessary to manage it. It is worth noting that the importance of files lies in the value of the data stored in them. When vulnerable to reviewing or modifying by intruders, this data affects companies and institutions in terms of security.

Cloud Computing is to use computer resources such as networks, servers, storage, applications, and services over the Internet while maintaining confidentiality, availability and integrity. The National Institute of Standards and Technology (NIST) defined various standard service models such as infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS) and software-as-a-service (SaaS). The IaaS enables customers to operate over operating systems, storage with limited control of network. The PaaS allows customers to deploy applications with some configuration of hosting applications. The (SaaS) supports customers to use applications on cloud with limited configuration settings. NIST defined standard deployment model such as the Private Model used by a single organization, the Community Model used by a specific group of institutions sharing common goals, the Public Model used by the public, and the Hybrid Model used by a composition of two or more cloud models (private-community-public)[1]. The said cloud models are believed to function better if they are protected in terms of confidentiality and privacy.

Cloud confidentiality and privacy is a very critical term in these days because of the huge usability of cloud services. Several applications and research papers handle the issue of cloud privacy. These can be categorized into two terms: fragmentation-defragmentation and encryption-decryption. Privacy is a core challenge in cloud computing with regard
to the need to protect identity information and policy components. Many organizations are not comfortable in storing their data and applications on systems that reside outside their own premises and data centers [2].

Cloud computing technologies provide organizations and individuals alike with the advantage to access their files from anywhere. Yet, they are becoming more risky as the number of intruders increases[1]. Thus, the growing dependency on cloud computing led to a rising concern regarding cloud privacy. Cloud privacy is intended to protect personal data from collection, usage, disclosure, storage, and damage. It is very essential to protect the data from unauthenticated access. Hence, it is necessary to identify an optimal solution to protect data on clouds. There are different techniques for cloud privacy such as Encryption and fragmentation.

The idea of cryptography depends on building methods and protocols to protect private messages to prevent any intruders from accessing them based on mathematical problems. It’s used in several types of information security including protecting data integrity and privacy.

Encryption is a mechanism to protect data or messages from unauthorized access to information by any unauthorized persons, and it’s used in many methods of data security such as protection of integrity of data or protection of privacy. There are several methods of encryption including symmetric (also known as Public-key cryptography). Symmetric algorithms have several types such as Data Encryption Standard, Advanced Encryption Standard, Twofish and Blowfish. However, encryption has defects such as difficulty in searching and modifying. [3]. On the other hand, using encryption alone makes full files in the hand. Therefore, these files may be decrypted and, thus, the privacy is violated.

Beside encryption, fragmentation is a valid technique for cloud privacy. The main objective of fragmentation is to segment data into several parts and in different ways and store them on different cloud servers. Fragmentation decreases processing time and optimizes data manipulation in terms of transferring data across clouds, to illustrate this, by experimenting with sending a group of small files is faster than sending a file with the same size of small files, depending on the Internet speed. In this context, several studies provided evidence for using fragmentation as a solution in tabular data including horizontal fragmentation, vertical fragmentation, and hybrid fragmentation[4] and [5].

Based on what has been said, This paper introduced a model that aspires to produce a highly reliable model for cloud privacy using a combination of text fragmentation and encryption.

The overarching objective is to develop an applicable model that good protecting the privacy of text files on the cloud environment by combining a well-known encryption algorithm with diagonal fragmentation techniques with good time. The value of this study lies in providing organizations such as businesses and academic institutions with workable solutions to address such a critical issue as textual cloud confidentiality. It makes important contribution of a useful textual cloud privacy model that combines an encryption algorithm with a diagonal fragmentation technique that organizations can use to control the textual cloud storage with high confidentiality while achieving better confidentiality to maximize their information security.

Our model avoids the problem of not owning a cloud environment of our own, in addition to our distrust in the owner of that environment. It is distinguished from other approaches in that the file will be fragmented after encrypting it to send every part to a different cloud. By doing so, it increases the strength of the encryption technique being used, as there will be no way to decrypt an incomplete file. In addition, this technique isolates every part in ways that prevent the cloud owner to assemble it.

II RELATED WORKS

It is believed that encryption alone cannot provide security that can be destroyed by assaults or brute force techniques. [6]Named as Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions; they used an encryption key that represents a combination of a user's password and file name and which is changed to bits to defend against brute force attack. The privacy of the file is secured by multiplication of matrix, and the validity is guaranteed by using a hash-based message. Attribute Based Encryption (ABE), known as fuzzy identity based or Secret key, is provided by a collection of attributes. They also described the form of access used to control access. This access control uses encryption to encrypt and share data between users. (Public key encryption is used to encrypt the data by using the public key. Only the one who has the private key can decrypt this data.

[7]Introduced a confidentiality technique that is based on integrating encryption of confusion. Confusion uses a function of mathematics or uses programming techniques to misinform illegal users. For numeric data type, the obfuscation algorithm is used. The Obfuscation is a technique that uses specific mathematical functions or programming techniques to confuse data. When the data is alphabet or alphanumeric, the data will be encrypted using Symmetric encryption because of its speed and computational efficiency to handle large data volume encryption.

[8] Suggested a solution to the privacy issue by saving the CPU and the memory using symmetric algorithm data encryption in mobile cloud computing, sending it to the private cloud, and then re-encrypting the data and sending it to the public cloud via asymmetric algorithm. The experimental results obtained from comparing and evaluating encryption algorithms such as symmetric blowfish algorithm and asymmetric DSA algorithm to get the least time to decrypt data.

[9] Proposed an algorithm that provides protection for confidentiality and integrity. The key for encryption is concatenated with a user's password. Coding based Scheme (CoS) is used in the situation of Uploading process where the system
forces the user to enter a password before uploading the file to the cloud. Also, in the case of downloading files, the system requires entering the correct password. Encryption based Scheme (EnS) includes MD that performs data authentication and integrity checks.

[10] Used a hybrid fragmentation by segmentation in vertical and horizontal on tabular data. In order to divide the data into several types, one type will be encrypted, another will be sent without encryption, and the third one will not need to be sent and should be stored on the owner's hand, all that were categorized in term of its sensitivity.

[11] Proposed a strategy of confidentiality based on hosting number of virtual machine (V1……Vn) then splitting the file into number of chunks (C1…..Cn) and sending every chunk randomly to its virtual storage. By fragmenting the file, the resulting information will not be complete on every virtual storage, thus making it difficult for intruders to violate the information.

[12] Presented a vertical fragmentation algorithm in the system with complex attributes and complex methods. With distributed Object Oriented Database Systems, this form of fragmentation allows query decomposition, optimization, and concurrent treatment.

[13] Applied a vertical partitioning algorithm application to the problem of horizontal fragmentation. The main idea was to guarantee confidentiality by developing application of a vertical partitioning algorithm to horizontal partitioning. They achieved successful results.

[8] Applied and tested many cryptographic algorithms such as DES, 3DES, AES, RSA and blowfish. It was concluded the blowfish algorithm documented the least encryption time whereas the RSA algorithm recorded the slowest encryption time. If confidentiality and integrity are important factors, then the AES algorithm should be selected.

It is concluded that there remains a problem in previous research regarding dealing with the protection of the privacy of text files using individual encryption because it keeps the file complete in hand on the cloud. We know that it is difficult to be certain on testing the privacy. Therefore, we intend to increase the complexity of confusion by utilizing the idea of fragmentation. Several approaches to protect the Cloud data.

## III MATERIAL AND METHODS

This section presents the methodology used to achieve the research objective, which represents the privacy of text files on the cloud environment. It demonstrates the proposed architecture to determine the most effective confidentiality techniques. It also clarifies the implementation process of the confidentiality protection approach. The implementation is based on the architecture of the approach and realizes the confidentiality techniques specified as a basis for the data encryption and fragmentation. The steps involved in developing the textual file confidentiality model are further elaborated below.

### A Methodology of the model

Java language was used in programming the model because it provides all the capabilities required for the research work.

The model was divided into many sequential steps as shown in the following subsections illustrated in Figures 1 for uploading the file.

Figure 1 shows the sequential steps involved in downloading the file including collecting the parts of the file from the cloud servers on which the file was previously uploaded and then defragmenting it to obtain the original file.

### III MATERIAL AND METHODS

This section presents the methodology used to achieve the research objective, which represents the privacy of text files on the cloud environment. It demonstrates the proposed architecture to determine the most effective confidentiality techniques. It also clarifies the implementation process of the confidentiality protection approach. The implementation is based on the architecture of the approach and realizes the confidentiality techniques specified as a basis for the data encryption and fragmentation. The steps involved in developing the textual file confidentiality model are further elaborated below.
Implement the Model

The experimental environment consists of an HP omen “15t-dc100” laptop with the following specifications: Intel(R) Core(TM) i7-9750 H VPU @2.60GHZ 2.59 GHz, 16G RAM and GeForce RTX 2070 GPU (8GB of VRAM). We Use Java Language in our system, SQL Database and Dropbox cloud storage files.

Secure the File Strategy

The main idea is to split the file diagonally into four parts (P1, P2, P3, P4) before uploading them to the cloud environment and then uploading each part separately on a different server (S1, S2, S3, S4), after encrypting it. This process makes it difficult for the intruder to access the data because the incomplete, splitted file cannot be decrypted.

Encryption

As shown in Figure 1, the model starts with browsing files and determining the file to be uploaded to the cloud environment. The second step (Step 2) relies on pre-uploading file encryption, evaluating, and selecting the best and fastest symmetric algorithm of encryption. According to [13], it's known that the symmetric encryption like Blowfish has a higher encoding speed than the asymmetric encryption like RSA.

Some key parts of the model were including checking the encryption mechanism of the Blowfish algorithm, as shown in the Figure 3 of the pseudocode; using Java language; and implementing a simple and successful experiment on a text file. The encryption process was executed using the Blowfish algorithm from a client side.

Fragmentation

As a third step (Step 3) in Figure 1, and for detailing this part. Two important factors must be considered when choosing a partitioning mechanism. The first is the accuracy ratio for privacy, and the second is the system speed. The figure 4 below shows the result of the fragmentation into two parts, and it extracted that error ratio approach in terms of privacy accuracy using the formula for calculating it, which is large, so we go to split the file into several chunks.

The file retrieval strategy

This is the reverse process of the previous steps explained above (Secure the file strategy). First, the file is selected for download through the control panel so that the model identifies the servers where the files are located as described in the first step (Step 1) in Figure 2. Then, the parts of the file are defragmented as described in the second step (Step 2) of the same figure. Later, the four files are compiled in a reverse manner of the separation process, and then the same key is used to decode the file. Then, the file is decrypted as shown in the third step (Step 3) of figure 2.

Experiments and Results

A practical experiments and results of the research to measure the effectiveness and the efficiency of the Diagonal Fragmentation and Encryption model in terms of high privacy for text files on the Cloud environment. In the experiments, the privacy Accuracy, the size of files, the time needed for the fragmentation and the encryption processes, and the time needed to retrieve the file were taken into consider-
Privacy accuracy refers to using the count of complete words 'CW' divided by the whole words count 'AW' in the files as demonstrated in the following equation:

Privacy Accuracy ‘PA’ = (1-(CW/AW))*100

Where CW is 'complete words' count and AW is 'all words' count.

The counting process takes place in two stages in the experiment. The first stage is to count all the words ‘AW’ in the original file and this is done programmatically before the fragmentation process. The second stage are done after fragmentation process to count completed words ‘CW’ stilled in the fragmented files and this is an error rate, and this process is done manually.

A Files
As shown in the Microsoft Windows Explorer Standard File Size Classification table below, various text files of different sizes were used in the experiments to check the sizes before and after fragmentation and to measure the time required for the fragmentation process. Accordingly, the experiments were conducted on small, medium, and large files.

| Microsoft Windows Explorer File Size Classification [22] |
|----------------------------------------------------------|
| **Small** | 10-100 KB |
| **Medium** | 100 KB-1 MB |
| **Large** | 1-16 MB |
| **Gigantic** | >128 MB |

B Effect of Fragmentation
We all know that to increase security you need a cost in terms of time and accuracy, so in our practical experience we will explain the amount of this cost. It must be clarified that to preserve the integrity of the data, all the fragmented files must be preserved to can recover the original file.

i DIAGONAL FRAGMENTATION FOR THE FILE AS ONE PART:
In this experiment, the file was divided diagonally as one chunk to take the necessary readings, including the percentage of privacy and the ratio of the impact of the division on the size of the files and the time needed for the fragmentation.

Figure 5 below shows the time needed to divide the 984-byte file diagonally into two blocks or 4 blocks as one chunk. The results show that the model took more time to divide into 4 blocks than to divide two blocks.

Fig. 5. Privacy Accuracy of small file using 4 -2 files

ii CHUNK: DIAGONAL FRAGMENTATION FOR THE FILE AS MULTIPLE PART
When examining files with a larger size, such as 27 KB, the percentage of privacy accuracy decreased. The privacy accuracy ‘PA’=(1-(1586/3817))*100=58.44% for the 27KB file as show in figure 6. Therefore, there is a need to develop a more advanced diagonal fragmentation for the text file by dividing the file to a number of chunks depending on the word count, and then dividing each chunk diagonally into two parts. The file, thus, was divided into chunks of words. Each chunk contains a fixed number of words, and then these chunks were divided diagonally into two parts.

Fig. 6. Privacy Accuracy of small file (27KB) using 4 files

iii Measuring the Best Value for a Chunk Word Count
In order to be able to measure the best value for word count in each chunk, there is a need to use two important factors for this purpose. The first is the time needed by the model to use this number of words, and the second is the privacy accuracy the model achieves using the same number of words.

Time consumption factor :Figure 7 shows the results for time consumption achieved by the model using three different words counts categories, namely, 100 words, 200 words, and 500 words in each chunk. The results show that the time of division by these three-word count categories was convergent for the three file size categories, illustrated in Table
iv Privacy accuracy factor:

Given the privacy accuracy equation (Privacy Accuracy ‘PA’ = (1-(CW/AW))*100), the model achieved accuracy levels of 96%, 90%, and 88% when using 100 and 200 words in each chunk on small, medium, and large files respectively. As showed in figure 8, the case with 500 words in each chunk was not optimal as the privacy accuracy percentage ranged from 76% to 82%, which are not accepted.

Fig. 7. Privacy Accuracy ‘PA’ = (1-(CW/AW))*100.

D Sending Data to the Cloud

The time needed to transfer the four files to the cloud was examined to compare it with the original single file transfer time without fragmentation to the cloud. The purpose was to measure the model efficiency in terms of fragmentation upload time.

The results demonstrated a time increase when dividing files and distributing them on four servers on the cloud as shown in the table 2. However, the upload time depended on two additional factors: the speed of the Internet and the cloud environment to which the files will be transferred. This issue can be resolved by increasing the speed of the internet and choosing a fast cloud service.

TABLE 2
Upload time for the Files to Cloud.

|                | One File transfer time | Four Files transfer time |
|----------------|------------------------|--------------------------|
| 3 KB           | 0.7 second             | 4 second                 |
| 3.429 KB       | 2.226 second           | 4 second                 |

V CONCLUSION AND FUTURE WORKS

This comprises the conclusion of the research. It is intended to highlight the major contribution of the research to the existing body of knowledge in information security, with particular emphasis on confidentiality. Also, it stresses the future work that needs to be attended by further research.

The major contribution of this research is the presenting of an effective text file privacy model on the cloud based on diagonal fragmentation and encryption. The synergy between encryption and fragmentation added strength and accuracy to the privacy of cloud textual files. The idea is that, encryption alone does not mean that the file is not fully secured because once the file is decrypted, then the privacy is violated. Yet, application of the encryption and the fragmentation technologies in one model makes it impossible to violate the privacy simply because an initially encrypted and then fragmented file will never make the information accessible to intruders even if they had the fragmented files in hand.

Another contribution of the current study is the application of fragmentation techniques to textual files.

Five experiments were carried out in this research with five different methods of fragmentation, each of which was performed on different sizes of textual files ranging from small, medium to large files in accordance with the standard classifications of file size. In each experiment, the time used by the model to secure the file on the cloud and then retrieve it was calculated. Besides, the privacy strength was also calculated using a formula that was developed to measure the strength of privacy.

Furthermore, this research introduced an equation (Privacy Accuracy ‘PA’ = (1-(CW/AW))*100), to measure privacy accuracy, thus resolving the issue of privacy measurement. It is worth noting that the model developed in this study achieved privacy accuracy exceeding 96% depending on fragmentation only. After integrating encryption with frag-
Text File Privacy on the Cloud Based on Diagonal Fragmentation and Encryption

Tawfiq Barhoom and Mahmoud Y. Abu Shawish

Abstract

In the cloud storage era, ensuring privacy and security of user data is of utmost importance. This paper introduces a novel approach for text file privacy on the cloud based on diagonal fragmentation and encryption. The approach involves dividing large textual files into smaller frames, applying a diagonal fragmentation method to achieve a high level of data fragmentation, and encrypting the data using Blowfish encryption. This technique not only enhances data security but also improves the efficiency of data retrieval and storage in cloud environments. The proposed method is validated through simulation and real-world cloud storage scenarios, demonstrating its effectiveness in maintaining privacy and security standards.

Keywords: Cloud storage, Text file privacy, Diagonal fragmentation, Blowfish encryption.

1. Introduction

In recent years, cloud computing has become a prevalent technology for storing and accessing user data. Despite its numerous benefits, such as scalability, cost-effectiveness, and accessibility, it also poses significant privacy and security challenges. Ensuring the privacy and security of user data is crucial, especially in the context of sensitive information like personal data and intellectual property.

This paper focuses on the problem of text file privacy on the cloud, aiming to develop an effective and efficient approach to protect data confidentiality. The proposed method combines diagonal fragmentation with Blowfish encryption to achieve a robust data protection mechanism.

2. Related Work

Several approaches have been proposed in the literature to address data privacy and security in cloud storage systems. Diagonal fragmentation (Do et al., 2017) has been employed to enhance data security by making it difficult to retrieve the original data from fragmented pieces. Blowfish encryption (Schneier, 1993) is a symmetric key block cipher known for its efficiency and security.

3. Proposed Approach

The proposed approach involves the following steps:

- **Diagonal Fragmentation:** Large textual files are divided into smaller frames using a diagonal fragmentation technique to ensure that even if some parts are compromised, the original data cannot be reconstructed.
- **Blowfish Encryption:** Each fragment is encrypted using the Blowfish encryption algorithm to prevent unauthorized access.

4. Implementation

The implementation of the proposed approach involves customizing a cloud storage service to integrate the diagonal fragmentation and Blowfish encryption modules. This requires modifying the storage and retrieval processes to accommodate the fragmented and encrypted data.

5. Evaluation

The effectiveness of the proposed approach is evaluated through simulations and real-world cloud storage scenarios. The results demonstrate a significant improvement in data confidentiality compared to existing methods.

6. Conclusion

This paper presents a novel approach for text file privacy on the cloud based on diagonal fragmentation and Blowfish encryption. The method offers a balance between data security and usability, making it a promising solution in the era of cloud storage.

References

[1] P. Mell and T. Grance, “The NIST-National Institute of Standards and Technology- Definition of Cloud Computing,” NIST Spec. Publ. 800-145, p. 7, 2011.
[2] D. Chen and H. Zhao, “Data security and privacy protection issues in cloud computing,” Proc. 2012 Int. Conf. Comput. Sci. Electron. Eng. ICCSEE 2012, vol. 1, no. 973, pp. 647–651, 2012.
[3] V. Mai and I. Khalil, “Design and implementation of a secure cloud-based billing model for smart meters as an Internet of things using homomorphic cryptography,” Futur. Gener. Comput. Syst., vol. 72, pp. 327–338, 2017.
[4] T. Kalidoss et al., “Data anonymisation of vertically partitioned data using Map Reduce techniques on cloud,” Int. J. Commun. Networks Distrib. Syst., vol. 20, no. 4, pp. 519–531, 2018.
[5] A. Hudic et al., “Data confidentiality using fragmentation in cloud computing,” Int. J. Pervasive Comput. Commun., vol. 9, no. 1, pp. 37–51, 2013.
[6] S. Aldossary and W. Allen, “Data Security, Privacy, Availability and Integrity in Cloud Computing: Issues and Current Solutions,” Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 4, 2016.
[7] L. Arockiam and S. Monikandan, “Efficient cloud storage confidentiality to ensure data security,” 2014 Int. Conf. Comput. Commun. Informatics Ushering Technol. Tomorrow, Today, ICCCI 2014, pp. 1–5, 2014.
[8] T. S. Barhoom and M. M. Abu Ghosh, “Reduce Resources for Privacy in Mobile Cloud Computing Using Blowfish and DSA Algorithms,” Int. J. Res. Eng. Sci. (IJRES) ISSN (Online), vol. 4, no. 1, pp. 2320–9364, 2016.
[9] W. Ren et al., “Lightweight and compromise resilient storage outsourcing with distributed secure accessibility in mobile cloud computing,” Tsinghua Sci. Technol., vol. 16, no. 5, pp. 520–528, 2011.
[10] R. Hussein Al-Talaa, “A Confidentiality Protection Approach Based on Three-Way Fragmentation for Cloud Outsourcing of Mobile Data,” 2015.
[11] A. Butoi and N. Tomai, “Secret sharing scheme for data confidentiality preserving in a public-private hybrid cloud storage approach,” Proc. 2014 IEEE/ACM 7th Int. Conf. Util. Cloud Comput. UCC 2014, no. see IV, pp. 992–997, 2014.

Tawfiq S. Barhoom
Associate Professor, Computer Science Department, Faculty of IT, Islamic University-Gaza, he got B.Sc. Computer Science from Omdurman Ahlia University (Sudan 1991-1995). Master degree and PhD Computer science, Department of computer science and engineering from Shang hai Jiao Tong University (SJTU)- China (1999, 2004 respectively). His current interest research information security.

Mahmoud Y. Abu Shawish
was born in Gaza City, Palestine, on the. The instructor at Information Engineering department, University College of Applied Sciences, Gaza, holder of BA. Computer engineering (2002-2007) and Mas-ters Degree in information technology from Islamic University, Pales-tine. His current interest research Cyber Security.