Protecting Data in Personal Cloud Storage with Security Classifications

Fara Yahya
Electronics and Computer Science,
University of Southampton, Southampton,
United Kingdom
fara.yahya@soton.ac.uk

Robert J Walters, Gary B Wills
Electronics and Computer Science,
University of Southampton, Southampton,
United Kingdom
rjw1, gbw @ecs.soton.ac.uk

***Abstract*— More and more data is stored in personal cloud storage and it is expected to grow further. As cloud storage is becoming an option for user in keeping their data online, it comes with the security threats and the challenges of protecting their data from unauthorised access. Many security controls have been implemented by cloud storage providers (CSPs) as a security measure but although encryption is known as one of the solution, it cannot be fully implemented due to the need of a robust and costly infrastructure. Therefore, we foresee the need to implement protection based on categorisation of data determined by users. This paper will discuss recent security threats and also the levels of data protection in cloud storage. We propose a security framework that protects data in cloud storage based on the level of protection it needs.**

Keywords*— *Cloud Storage, Cloud Storage Provider, Security Classification

I. INTRODUCTION

Personal cloud storage provides the facility for user requiring mainly highly scalable storage on demand and accessible globally. Commercial leading personal cloud storage includes Dropbox, Box and Google drive etc. have been used as a medium to store personal data in the cloud as shown in figure 1. According to Gartner, user will be storing more than 36% of their data in the cloud by 2016 [1]. This is projected to be 3.3 terabytes for each user. Despite the benefits it brings to user in particular, in terms of convenience and lower cost, it also brings security concerns. These concerns have been supported by various researches [2]–[6]; security threats are increasing significantly every year.

In recent reports by CSA and Georgia Tech Information Security Center (GTISC), data loss and leakage threats are immensely discussed as outsiders are gaining access to unencrypted data. This can simply be undertaken using several methods: a user passwords can be cracked or brute-forced, and then unencrypted local file/s or folder/s located in the cloud can easily be accessed. On the other hand, the cloud service itself can also be compromised. In these cases, users and CSPs can implement security measures before the data is exported online. Figure 1 below shows some of the current personal cloud storage in the market.

CSPs have been implementing controls to secure access to

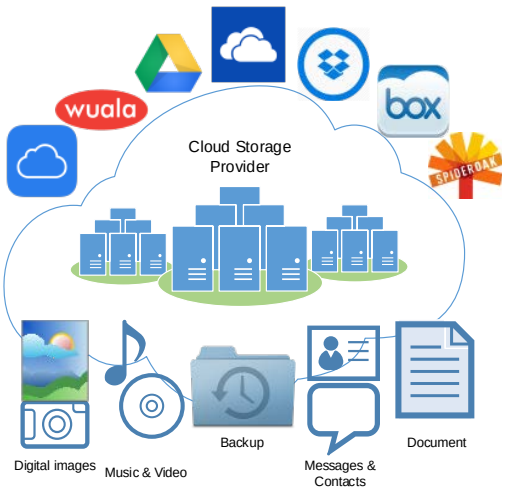

Figure 1 Personal cloud storage

sensitive data in the cloud such as two-factor authentication, encryption etc. making access to the data more difficult for attackers. Encryption is an effective and widely known as the primary solution to protect data but it is not fool-proof. Additionally, to encrypt the whole data on cloud in order to protect against unauthorised accessed will need a robust infrastructure and is greatly expensive to be enforced. Therefore, it has not been consider as the best option for CSPs.

Then again, an increase in security measures effects the usability of the data and therefore causing the system to be shunned by users. It is known that not all data stored in a cloud storage is private or confidential. Some of the data is less important and therefore need basic protection. Most CSPs are unwilling to reduce the efficiency of accessing into cloud storages because users expect an equally efficient access into a secured data as the plain text ones. We foresee the effort of protection based on an acknowledged security level of data determined by the users.

Security levels for data protection can be applied as an option to protect data in cloud storage. There are various ways of protecting a data such as categorising it into several security

groups having different level of protection mechanism. For an example, in the military services, several categorisations are used to protect assets such as top secret, secret, and official. For each category, different level of protection is applied. Imagine top secret assets are protected in-depth with multi-layer of shields before the asset can be accessed. Another example would be the concept of safety box in some banks; it protects assets in a locked storage for valuable items but the cost is charged to the user.

This paper will be arranged in several sections. The introduction will give a brief idea of the proposed research. The next section will look into security threats in cloud storage and later a security level of data protection against the threats. The related work will discuss previous research related to security challenges in cloud storage. A section that will explain an overview of the proposed framework is also done followed by conclusion and future work.

II. Security Threats

Cloud storage is a service that comprises of benefits and also challenges. It inherent vulnerabilities, but these have never discourage users from taking advantage of its functionality and flexibilities. Cloud users are data owners that have concerns whether their data are secured and protected in the cloud. With the adoption of a cloud model, users lose control over data security. In fact, in most known cloud storage, users are sharing the resources with other users.

Security threats is a possible vulnerability that may breach security and cause harm to a user or organisation. These threats are potential in causing adverse impact. A threat may be happening from inside or outside of an organisation or either intentional or accidental. In previous researches, it is shown many security threats are happening in the cloud. We will review security threats happening in a cloud storage in this section [2]–[6].

Table 1 Security Threats in the Cloud

Threats	Cloud Security Alliance (2010, 2013)	Georgia Institute of Technology (2013)	Sabahi (2011)	Bashir & Haider (2011)
Password cracking or Brute Force	Severe	-	Low	Severe
Inconsistent Use of Encryption	Severe	Severe	Severe	Moderate
Catastrophic Hardware Failure	Moderate	-	-	Low
Malware	Low	Severe	Severe	Severe
DDoS	Low	-	Low	-
Man in the middle attack	-	Moderate	-	Low

In a statistical overview of vulnerabilities report by the Cloud Security Alliance (CSA) mentioned that the highest incident occurred from threats as followed; insecure Application Programming Interface/s (APIs), followed closely by data loss and leakage and thirdly, hardware failure from twelve threats defined by CSA [2]. Among threats involved in insecure APIs are anonymous access and/or reusable tokens or passwords, clear-text authentication, improper authorisations or API dependencies [3]. These vulnerabilities contributes to password cracking or brute force attacks. On the other hand, unauthorised access must be prevented from sensitive data which involves inconsistent use of encryption and software keys. Lastly, a CSP that suffered from a catastrophic hardware failure denying user access to their data.

A report done by the Georgia Institute of Technology on the emerging cyber threats. According to their report [4], the highlight was on data security on the cloud allowing unauthorised access to unencrypted data. In the cloud, user needs to protect information from data-stealing malware as cybercriminals has been using this technique to access data using trusted and reputable cloud storages services from which malware can be downloaded. Once a malware has infected the user device, it will be hard to identify. Although malware bring lesser threats to mobile user but man-in-the-middle attacks are increasing as more users are connected on untrusted networks using their smartphones and tablets. Eventually they will access cloud storage services from these devices too.

Sabahi [5] in his paper emphasis on the reliability, availability and security in the cloud. The data is moving inside the cloud as it is a multi-tenant environment and this raise an important concern called data leakage. An unauthorised user may launch an attack using malicious codes in virtual machines and if a hacker take control over it, sooner or later being able to access data within it. Therefore, unencrypted data can be copied and used. There were also issues on attacks like Distributed Denial of Service (DDoS) attacks of having much greater impact on multi-tenant architecture. When a DDoS attack is launched, it floods a packets to a Web server from multiple sources making the service unavailable resulting in inaccessibility.

Another research on security threats by Bashir and Haider (2011) has focus on addressing security concerns in the cloud [6]. The study ranked security as the primary challenge in cloud. Among discussed is the data centric cloud security by securing data query processing. Data must be protected specifically during processing using techniques such as encryption to ensure it is not viewed by the computer administrator. Also, filtering of the system as the major step in finding malicious codes planted. Some other concerns involve security breaches due to the weakness of security policy such as password complexity and tokens. Sensitive information such as passwords must be hardened or it will easily be cracked by attackers.

III. Optimising Data Classification in Determining Security Levels of Protection

The cloud is a multi-tenant environment, where resources are shared. Threats can happen from anywhere; inside the shared

environment or from outside of it. However, placing sensitive data in a shared cloud storage is apparently risky. Whether accidental or due to a malicious hacker attack, data privacy, loss or leakage and unavailable for access would be a major security violation involving confidentiality, integrity and availability.

The best strategy is to secure based on protection levels in cloud storage. An interview report done to three Chief Information Security Officers (CISO) of Royal Bank of Scotland, DELL corporation and Microsoft mentioned that, if organisation can apply the right level of control for assuring its data confidentiality, it will deliver a significant benefit in managing their security protection [7]. It is known that not all data stored in a cloud storage are private and confidential. Some are less important and therefore need basic protection. Most CSPs are unwilling to reduce the efficiency of accessing into cloud storages because users expect an equally efficient access into a secured data as the plain text ones. We are emphasising on protection based on an acknowledged security classification of data determined by the users.

There are various ways of protecting a data such as categorising it into several security groups having different level of protection mechanism as shown in figure 2 below.

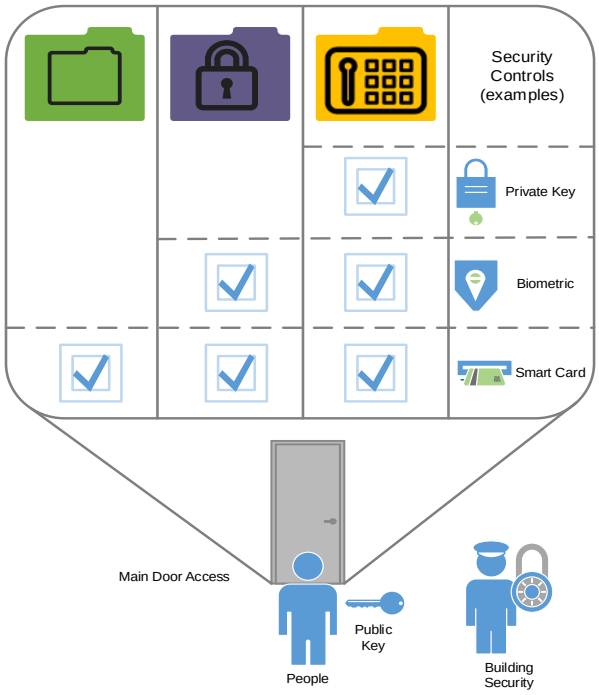

Figure 2 Examples in a Protection Levels

Security classification is known as the process of managing and organising security protection into levels and categories for its most effective and efficient application. A well-planned security classification system makes data protection easier to implement. This can be of particular importance for risk management, legal discovery, and compliance. Assigning a security level of protection to different data classification in cloud storage will give different level of sensitivity to classified information.

Users are able to manage their data protection by having assign a value based on the level of sensitivity. An effective security classification involves a broad awareness of users understanding of data residing in a cloud storage [8]. Data exists in one of three basic states: at rest, in process, and in transit. All three states require unique security solutions for data protection, but the applied principles of security classification should be the same for each. For an example, data that is classified as sensitive needs to stay sensitive when at rest, in process, and in transit.

Data can also be either structured or unstructured. Typical classification processes for structured data found in databases and spreadsheets are less complex and time-consuming to manage than those for unstructured data such as documents, source code, and email. Generally, users will have more unstructured data than structured data. Regardless of whether data is structured or unstructured, it is important for users to manage the level of sensitivity. Therefore when properly implemented, security classification helps ensure that sensitive or classified data are managed with greater oversight than data that are considered public or free to distribute. By having all information on authentication, authorisation and encryption will assist user in understanding whether the cloud storage provider supports the data protection requirements mandated by their security classification as below:

A. Authentication

Authentication typically consists of at least two parts: a username or user ID to identify a user and a token, such as a password, to confirm that the username credential is valid. The process does not provide the authenticated user with access to any items or services; it verifies that the user is who they say they are.

B. Authorisation

Authorisation provides an authenticated user the ability to access an application, data set, data file, or some other object. Assigning authenticated users the rights to use, modify, or delete items that they can access requires attention to data classification. Successful authorisation requires implementation of a mechanism to validate individual users needs to access files and information based on a combination of role, security policy, and risk policy considerations. In ensuring access controls to who can see which and when there must be an effective authenticating system in place.

C. Encryption

Encryption has always been seen as the ultimate security measure to protect data at rest, in process, and in transit. Retaining encryption keys have also been a concern for users storing their data in cloud storage.

Further discussion on related work in securing access to a cloud storage particularly; authentication, authorisation, and encryption is presented in the next section. This will give an overview of previous researches done focusing in this area.

IV. RELATED WORK

Previous researches on cloud storage have emphasis on a wide range of technical approaches over the aforementioned concerns. Access security measures are generally considered in three steps: Authentication, Authorisation and Encryption. Some security measure includes effort to secure access based on hardening passwords [9]–[11]. Generating strong passwords and protecting them from getting stolen guarantees a password security. Researchers have established that strong passwords are necessarily long, random and hard to crack but often difficult to remember. Bang et al. suggests that security is not just a technical issue but also a behavioural issue involving users, mostly untrained ones [12].

An authorisation process ensures that a person has the right to assess a certain re-sources and limits of the access unknowing of other user information. Users may have access but have a specific role or authority to do something within their scope. A paper suggested an authorisation model suitable for cloud services that supports hierarchical role-based access control (RBAC), path-based object hierarchies and federation [13] in multi-tenancy environment. These features provide a convenient authorisation service for cloud, especially those using path-based patterns such as REST APIs. Although authorisation usually supports high scalability, it is believed to improve scalability and this would hopefully enable more fine-grained control on the authorisation information.

A comprehensive approach using encryption ranging from data-in-transit to data-at-rest have been researched widely. Mostly developed a cryptographic cloud storage system; symmetric [14], [15] and asymmetric [16]–[19]. It is a standard approach to apply encryption techniques into sensitive data to secure it. Encryption has always been seen as the ultimate security measure but it also comes with a set of difficulties. Traditional encryption is done by transferring the data files locally and decrypting it. A cryptographic cloud storage system called CS2 was amongst early research done on applying symmetric encryption techniques that ensures confidentiality, integrity and verifiability without being resource hungry [14].

A Cloud storage encryption (CSE) framework was proposed also using a symmetric, searchable encryption with policy and access methods [15]. An Attribute-based encryption (ABE) with verification and recovery technique was proposed to effectively secure the data and provide recovery mechanism [19]. A different paper suggested ABE encryption was an efficient data retrieval scheme best suited for cloud storage systems with massive amount of data [18]. A fine-grained and cryptographic access control for cloud storage services called CS-CACS uses CP-ABE which is implemented based on the Hadoop Distributed File System (HDFS) environment [17] to efficiently secure user data. An approach was introduced [16] using user-centric privacy preserving cryptographic access control protocol, K2C (Key To Cloud) that enables end-users to store and share sensitive data securely in untrusted cloud storage for hierarchically organised data. It uses two cryptographic libraries, Hierarchical Identity-Based Encryption and Key-Policy Attribute-Based Encryption.

V. CONCEPTUAL CLOUD STORAGE SECURITY FRAMEWORK

In this paper, we propose a framework focusing on a varied level of security based on the category of data it falls into. We are looking into security protection for data stored in a cloud storage system, based on security classifications levels. This will also enable the implementation to be more flexible as it will offer multi-layers of security only to a group of data that requires it. In the effort of protecting data stored in a cloud storage, it is classified into different level of security classifications.

Discovering and implementing ways to protect data becomes an integral deployment strategy once a data has security classifications. Protecting classified data needs further consideration in terms of ways data is stored and transferred in existing architectures as well as in the cloud particularly cloud storage.

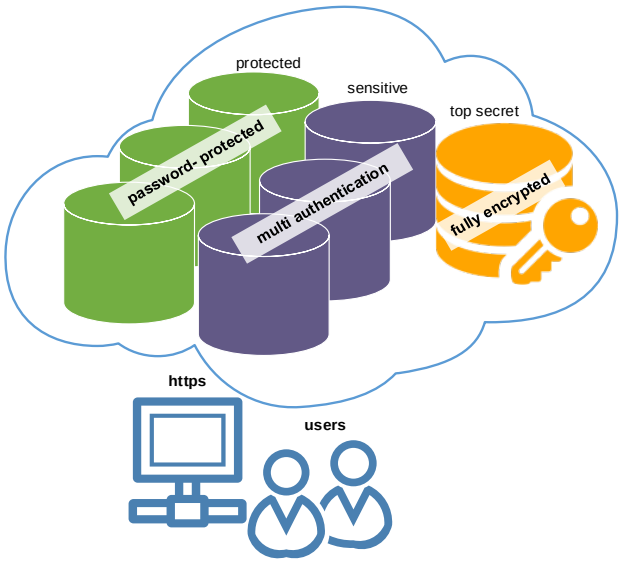

Figure 3 Security Classifications to Protect Data in Cloud Storage

As shown in figure 3 above, there are several security classification levels: protected, sensitive and top secret. Initially, a protected level involves data that is password protected as offered by most cloud storages. A sensitive level may need extra protection such as having multi-factor authentication. Some CSPs has introduced this as a security measure such as Dropbox and Google Drive. Finally, a top secret level may need to be fully encrypted even by the CSPs System Administrators. These security levels has different security weightage as some data needs less protection while some needs best protection. Therefore, in effort of securing access to cloud storage we focus on looking into the authentication, authorisation and encryption of current offered security solution by CSPs. In the next section, we will discuss the three levels of security classifications levels with existing

security technical solutions and recommended extra protections.

A. Security Protection Levels

In this framework, we propose three levels of security classifications: protected, sensitive and top secret. In table 2 below, the security protection levels in cloud storage is briefly shown. These security protections for protected and sensitive levels are based on existing control and measure by some known cloud storage providers.

Table 2 Security Protection Levels in Cloud Storage

Security Levels	Authentication	Authorisation	Encryption
Protected	Single Factor	Administrator	- SSL - 256-bit AES
Sensitive	Multi-Factor	- Administrator - Secure Data Access Sharing	- SSL - 256-bit AES
Top Secret	Multi-Factor	- Super Admin - Secure Data Access Sharing	- SSL - 256-bit AES - RSA - Filename Encryption

a) Protected

Protected level involves security protection for data that is for public or free distribution. Usually this includes data and that are not critical to user needs. This classification can also include data that has deliberately been shared to the public for use, such as marketing material. This level of protection is provided by most cloud storage provider in the market. Below are the main security characteristics:

- Single factor authentication

Single factor authentication usually involves single layer of security access such as password protected.

- Authorisation

A user are usually Administrators for their own data on cloud storage with privileges to create, edit and delete it.

- Encrypted at rest and in transit

A normal encryption method in a cloud storage involve protecting data in transit using Secure Sockets Layer (SSL)/Transport Layer Security (TLS) for data transfer to create a secure tunnel protected by 128-bit or higher Advanced Encryption Standard (AES) encryption. Once it reaches the cloud storage, it is protected using 256-bit AES encryption at rest.

b) Sensitive

Sensitive level involves security protection for data that is classified as being of medium sensitivity including data that would not have a severe impact on the user if lost or destroyed. Generally, this classification includes data for non-public view. This classification may include corporate data as most data that are accessed frequently or in daily use can be classified as sensitive. Below are the main security characteristics:

- Multi factor authentication

Multi factor authentication such as two-step verification or reentering password. Some CSP has introduced password protection as the first layer of authentication and another security codes sent to the registered mobile number or using a mobile app as the second authentication.

- Authorisation

A user are usually Administrators for their own data on cloud storage with privileges to create, edit and delete it.

- Encrypted at rest and in transit

A normal encryption method in a cloud storage involve protecting data in transit using Secure Sockets Layer (SSL)/Transport Layer Security (TLS) for data transfer to create a secure tunnel protected by 128-bit or higher Advanced Encryption Standard (AES) encryption. Once it reaches the cloud storage, it is protected using 256-bit AES encryption at rest.

c) Top Secret

Top secret level involves security protection for data that is classified as confidential or restricted including data that can be catastrophic to one or more user if com-promised or lost such as personal data, including personally identifiable information such as Social Security or national identification numbers (passport numbers etc.), specific intellectual property, legal data, authentication data (private cryptography keys, username password pairs, or other identification sequences such as private biometric key files). Below are the main security characteristics:

- Multi factor authentication

Multi factor authentication such as two-step verification or reentering password. Some CSP has introduced security codes sent to the registered mobile number or using a mobile app.

- Authorisation

In a top secret level, a user is a Super Admin with privileges to create, edit and delete data and but with highest level of access.

- Encrypted at rest, in process, and in transit

A top secret encryption method in a cloud storage involve protecting data in transit using Secure Sockets Layer (SSL)/Transport Layer Security (TLS) for data transfer to create a secure tunnel protected by 128-bit or higher Advanced Encryption Standard (AES) encryption. Once it reaches the cloud storage, it is protected using 256-bit AES encryption at rest. The data in process (in-use) is protected using 256-bit AES and RSA encryption. Data are encrypted locally before uploaded to cloud storage and the private key is kept by the user.

VI. Conclusion and Future Work

The cloud is a mutual environment, where users are sharing the resources to store their data online. Security threats are happening widely in the cloud. The threats includes, password cracking, inconsistent use of encryption, malware, hardware failure, DDoS, and Man in the middle attack. CSPs has introduced obligatory security measure and controls in undertaking these threats.

Although there are many security controls built-in to protect data stored in a cloud storage but a reliable framework that have security classifications for data stored in cloud storage has less been explored yet. Some solutions like total encryption is known as one of the appealing solution but it is barely implemented due to the need of a robust and costly infrastructure. Therefore, we propose a cloud storage security framework whereby the measure and controls are done based on security classifications.

Generally, security classifications can yield significant benefits, such as compliance efficiencies, improved ways to manage the security of users resources, and facilitation of data management in cloud. Although security classification efforts can be a difficult undertaking and require user assessment for successful implementation, quicker and simpler efforts can also bring benefits. Any security classification efforts should endeavour to understand the needs of each user and user can be more aware on data storing, processing capabilities, and data transmission in the cloud. The suggested security classification of protection levels: protected, sensitive and top secret are worth noting as a recommended security classifications guide. It is also expected to help reduce and mitigate risk with the suggested technical security solutions.

The future work of this research will investigate the data protection levels that can be applied by CSPs in a cloud storage architecture. This involve investigating security control and measures implemented by known CSPs. The adopted security model and its security classifications (if any) will also be explored and assessed as the current technology and guideline.

This will later involve a development of the security framework that can be used in a cloud storage architecture addressing the identified security classification for protecting data in cloud storage. The next phase will involve validating the idea with experts and simulations of the designated security classifications on cloud storage that will be used as the proof of concept.

Acknowledgment

We acknowledge the award of Malaysian Public Service Department Training (HLP) scholarship to Fara Yahya allowing the research to be undertaken.

References

[1] Gartner, “Newsroom: Consumers Will Store More Than a Third of Their Digital Content in the Cloud by 2016,” *Press Release*, 2012.

[2] CSA, “Top Threats to Cloud Computing V1.0,” 2010.

[3] CSA, “Cloud Computing Vulnerability Incidents : A Statistical Overview,” 2013.

[4] GTISC and GTRI, “Emerging Cyber Threats Report 2014,” 2013.

[5] F. Sabahi, “Cloud computing security threats and responses,” *2011 IEEE 3rd Int. Conf. Commun. Softw. Networks*, pp. 245–249, May 2011.

[6] F. Bashir Shaikh and S. Haider, “Security Threats in Cloud Computing,” *6th Int. Conf. Internet Technol. Secur. Trans. Abu Dhabi, UAE*, no. December, pp. 11–14, 2011.

[7] Microsoft, “CISO Perspectives: Data classification,” *Microsoft Trust. Comput. Doc.*, pp. 1–5, 2014.

[8] Frank Simorjay, “Data classification for cloud readiness,” *Microsoft Trust. Comput. Doc.*, pp. 1 – 19, 2014.

[9] M. Weir, S. Aggarwal, B. De Medeiros, and B. Glodek, “Password cracking using probabilistic context-free grammars,” in *Proceedings - IEEE Symposium on Security and Privacy*, 2009, pp. 391–405.

[10] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. López, “Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms,” in *Proceedings - IEEE Symposium on Security and Privacy*, 2012, pp. 523–537.

[11] R. Zhao and C. Yue, “Toward a secure and usable cloud-based password manager for web browsers,” *Comput. Secur.*, vol. 46, pp. 32–47, Oct. 2014.

[12] Y. Bang, D.-J. Lee, Y.-S. Bae, and J.-H. Ahn, “Improving information security management: An analysis of ID–password usage and a new login vulnerability measure,” *Int. J. Inf. Manage.*, vol. 32, no. 5, pp. 409–418, Oct. 2012.

[13] J. M. A. Calero, N. Edwards, J. Kirschnik, L. Wilcock, and M. Wray, “Toward a Multi-Tenancy Authorization System for Cloud Services,” no. December, 2010.

[14] S. Kamara, C. Papamanthou, and T. Roeder, “CS2 : A Searchable Cryptographic Cloud Storage System,” pp. 1–25, 2011.

[15] H. M. Al-sabri and S. M. Al-saleem, “Building a Cloud Storage Encryption (CSE) Architecture for Enhancing Cloud Security,” vol. 10, no. 2, pp. 259–266, 2013.

[16] S. Zarandioon, D. Yao, and V. Ganapathy, “K2C: Cryptographic cloud storage with lazy revocation and anonymous access,” in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering*, 2012, vol. 96 LNICST, pp. 59–76.

[17] R. Zhang and P. Chen, “A Dynamic Cryptographic Access Control Scheme in Cloud Storage Services,” *Int. J. Inf. Process. Manag.*, vol. 4, no. 1, pp. 104–111, Jan. 2013.

[18] D. Koo, J. Hur, and H. Yoon, “Secure and efficient data retrieval over encrypted data using attribute-based encryption in cloud storage,” *Comput. Electr. Eng.*, vol. 39, no. 1, pp. 34–46, Jan. 2013.

[19] R. V Agalya and K. K. Lekshmi, “A Verifiable Cloud Storage using Attribute Based Encryption and Outsourced Decryption with Recoverability,” vol. 3, no. 10, 2014.