Emerging Issues in Cloud Storage Security: Encryption, Key Management, Data Redundancy, Trust Mechanism

Daniel W.K. Tse, Danqing Chen, Qingshu Liu, Fan Wang, and Zhaoyi Wei

City University of Hong Kong
iswktse@cityu.edu.hk

Abstract. Cloud computing is one of the most cutting-edge advanced technologies around the world. According to the recent research, many CIOs who work for famous corporation mentioned that security issues have been the most critical obstacles in the adoption of cloud technology. As one of the prominent application of cloud computing, cloud storage has attracted more concerns; however, many security problems existing in cloud storage need to be resolved. This applied research paper briefly analyzes the development of cloud storage security, comprehensive discussion of several general solutions to those problems introducing several non-technical issues, such as third-party issues, trust mechanism in cloud computing. Finally, some possible improvement to those solutions is provided.

Keywords: Cloud storage, security, encryption, data redundancy, key management, trust mechanism**.**

1 Introduction

Cloud computing has provided enormous flexibility of its computation power to numerous users, including individuals, corporations and governments. The benefits brought by cloud services are obvious in terms of time and cost effectiveness. It is widely accepted that the cloud service will be a utility service after electrical utility services, water and gas services and telecommunication infrastructures.

Cloud storage is the cornerstone of cloud computing. Data security in cloud includes storage security, transmission security and processing security while storage security is a critical component. Gartner's survey indicates that the security problems have been the greatest obstacles on the way to cloud. From our intensive literature review exercise, most past research work done for cloud storage security is either too superficial or too theoretical. Because of this, we found the need to explore the nature of the problems by dissecting all critical scenarios and solutions comprehensively. This paper describes the loss of control in confidentiality, integrity and availability and then provides the solution to earn trust. Hopefully, some useful and high-impact improvements can be derived.

2 Research Methodology

Since the purpose of this research is to dissect traditional approaches to address the security issues in cloud and analyze some emerging techniques such as encryption,

L.S.-L. Wang et al. (Eds.): MISNC 2014, CCIS 473, pp. 297–310, 2014.

key management, trust mechanism and data distribution, these are very practical and complicated by nature, the traditional research methodologies are not appropriate to cope with such fast-technology world. Our research starts from some empirical studies and then drill down into the inner-working or mechanism of threats and their solutions. Thus, an applied research methodology, which is more practical and agile but weak in traditional research formality, is adopted.

3 Root Problems Analysis

3.1 Trust

Trust is a very important research area in cloud computing, especially in the field of cloud storage security. In traditional architectures, trust was enforced by an efficient security policy, which addressed constraints on internal control, constraints on access by external systems and adversaries including programs and access to data by people. In cloud architecture, this perception is totally obscured. Using public clouds, control is delegated to the organization owning the infrastructure to enforce a sufficient security policy which guarantees that appropriate security activities are being performed to ensure that risks have been reduced. All the policies and processes are behind the wire, users may think they are losing the control of their data.

3.2 Confidentiality

Confidentiality refers to data only being accessed by authorized people. Firstly, using the cloud to store data means delegating the authorization issue to cloud. Protecting user account from not being stolen and provide strong authentication can just lead to an external security. Avoiding internal threat is important too. The data in the cloud is facing the threats.

3.3 Integrity

Integrity is a key part of electronic information which means data can just be modified by authorized people. Within the local infrastructure, it is easy to limit and monitor the access of data so that integrity can be guaranteed. However, cloud is owned by other party, all the data uploaded to cloud is recorded and controlled by cloud service provider, users need a method to get control of the data's integrity.

3.4 Availability

Users want to get what they need when they want. Even though the requirement is not being satisfied, they have to know what is happening, why they cannot access, how and when it will be fixed. Using cloud, it seems that availability is never being a problem because cloud service provider has more server than you, so that they can perform backup and can hire professional person to manage the data center. As the

same situation stated before, the backup mechanism is processed in the cloud and user can never know whether there is a backup or not. It also makes user feel like loss of control.

4 Problems Analysis and Existing Solutions Discussion

4.1 The Basic Technology of Data Security -- Encryption

Encryption is a traditional and fundamental method to protect data. In this cloud era, it is still the first choice to safeguard the sensitive information.

In a multi-tenant environment [1], one of the basic security tools to protect data is encryption. When cloud service users do not have full control over the power in the cloud computing environment, cloud computing encryption allows users to protect the data effectively. This is extremely valuable for both private cloud and public cloud computing, especially when users share with other users a server which contains different levels of sensitive data repository.

When using cloud computing for data storage, usually virtual private storage architecture [2] should be used. Before sending the data to the cloud computing, encrypting the data, and decrypting them when they are sent back. For example, when cloud computing is a backup service, before storing the data in cloud, the backup software will use a local secret key to encrypt data in local. Because it is the users themselves who are in charge of encrypting operation and the secret key, keeping the key in the form of safety backup copy is one of the tools for data protection.

Volume label can be a useful medium to finish the encryption [2]. For encrypting operation to the data which are stored in IaaS application, the users can use equipment's volume label (volume label actually is not a file, but only a directory entry, it does not use extra disk space) to finish the encryption, and save the data in the second encrypted volume label. Due to the secret key and the encryption engine are all stored in local equipment, this is not the safest method but this method can protect the data from unauthorized access effectively. For example, assume that the encrypted volume label has been created correctly, the cloud computing providers will not be authorized to operate systems or applications (this is a typical default setting). Also, they cannot get the secret key and gain access to the encrypted volume label.

As more and more concerns are focusing on the encryption for cloud computing, there discovered several senior ways to improve the encryption method, one of them is using the third party broker [3]. We apply a third party broker rather than the users or cloud computing services providers to finish the encryption and decryption. The broker encrypts the clients' information, partitions the information into multiple segments and transfers them to corresponding virtual machines (VMs) of the cloud storage providers. The client information's integrity is ensured by a trusted third party called a cloud broker and it provides a more time-saving algorithm to encrypt, decrypt and detect the information.

Though encryption is significant for the security of implementing cloud computing, we should not treat it as the amulet because several problems [4] still need to be solved by the cloud computing services provider to ensure the clients' information

security. Firstly, protection during the encrypting: Services providers always encrypt sensitive data running in the cloud but if the data is not encrypted and being used or stored at the same time, reducing or preventing data from being destroyed will be their duty. Secondly, management of secret key: More and more internet users take part in the cloud services, along with the growing of the user group. The number of secret keys keeps increasing; it is a big challenge for services providers to protect data and a large number of secret keys. Thirdly, access control: Sometimes safety means complication. When layers and layers are added to the users data, it becomes not so convenient to get the data as the original intention of cloud storage. To ensure the user convenience so that user can use the data anytime anywhere they need in the safe way, it is necessary to build up an access control system.

For protection during the encryption process, there is another traditional way to realize the protection to the data being used - firewall. Deploying firewall which is dedicated for that kind of data, block the illegal access to both server side and client side. For secret key management, service providers should formulate severe management flow paths and rules and provide the secret key management system based on the hardware, like using independent encryption chips just for secret keys. For access control, the clients should require the encryption provider to give them enough power to access and management control and stronger authentication, such as two-factor authentication, access management and separate duties of safety management, such as security, networking, and operation maintenance.

4.2 Third Party Key Management Software

Cloud storage service is provided by big private company and it is self-supervised, so there is a high risk to have a policy fault or inner fraud in the company which can cause the data leakage. If the pernicious data is sent to the user's competitor or someone else who is hostile to this user, it will create a catastrophic damage. Obviously, the solution is to encrypt the data, enhance the difficulty for unauthorized people to get the useful information. There comes another question: How to transfer the key without the untrusted cloud? Using another cloud is a good idea. We need to make it more automatic, process the encryption and key storage in the background. It is a bridge between cloud and cloud user, called Key Management Software (KMS). KMS should be developed by professional companies which know about encryption, control the integrity checking, key generation, key distribution, periodic key changing and key destroy. Besides, KMS has to be verified by cloud service provider. The perfect mutual restriction is: Cloud takes charge of storage and the KMS takes charge of key management.

With such a restriction, cloud just needs to concern about the storage and lost recovery, the encryption efficiency and encryption deployment are transferred to KMS. It gives the chance for security companies to provide their service in a competition relationship which makes the cloud data decryption more difficult. For normal user, who just wants to share data and does not care about the confidentiality of data, he can use cloud storage service straightly. For small business companies who have the security concern, they can select which KMS they want according to their

budget and security requirement. For big company, they can even develop the KMS themselves.

Furthermore, KMS can do more than key management. As the storage and key management are separated, which means KMS and cloud are not with one-to-one relationship. A user can use different KMS to do different encryption based on the files security level. More importantly, the KMS can manage multi-cloud to provide an ultimate storage scheme: (1) Optimizing the download speed - Based on the network condition and server location, one cloud service provider cannot supply the best service anytime anywhere. But if user is using multi-cloud, he can test immediately and then find out the best performance he can get; (2) Never lose data - Cloud look like having infinite space for storage but users can never know whether their data has been backed up or not. At the worst situation, what if the company bankrupt? Not everyone can bear the risk of losing data. The best way to make user trust with is to store data in different clouds. This can be done by KMS and also KMS can check the data periodically; (3) Data splitting - This means cutting data into several pieces and storing them in different place, making bad guy hard to collect the entire data.

4.3 Distributive Storage in Cloud

4.3.1 Availability Threats

Availability means that the data stored in the remote cloud is available whenever the users request for it. There are many factors that may compromise data availability: Lack of replication and failure of servers, etc.

There are several kinds of risks related to cloud storage, like flood attacks (including direct DOS and indirect DOS) [5], fraudulent resource consumption attack, Single Point of Failure (SPOF) and the single point compromise, etc. Flood attacks, fraudulent resource consumption attacks and SPOF would harm the availability of data. Single point compromise would harm the confidentiality of data. Different data distribution scheme has different impact on both two aspects.

4.3.2 Defense Strategy: Redundancy Distribution Scheme

Redundancy

According to [6], in the traditional storage scheme, only one copy of the data is stored in the cloud. Once the only copy of data is corrupted or the server fails, the data will be temporarily or permanently unavailable to users. No data redundancy can meet the needs of data availability nowadays. The redundancy distribution scheme uses different techniques to guarantee that the data is available all the times.

For example, three clients store their data in the CSP1, CSP2 and CSP3 respectively. When CSP1 fails to function due to power-off or other reasons, client 1 cannot retrieve the data stored in CSP1. What's more disastrous is that if CSP1 cannot be repaired, the data stored in it will be gone forever. So no data redundancy can do great harm to both users and CSP. Many people turn to an alternative for help: Data is broken into several blocks and every block is stored in different servers. This approach is vulnerable to collusions among cloud service providers, which means the

providers holding different data blocks of the user can collude to reconstruct the original data. Consequently, encryption is still of great importance in distributed storage.

4.3.3 Different Distribution Schemes

Algorithm Comparison

A formula can be used to describe how different schemes influence the availability of data. The data distribution threshold algorithm [7] provides a comparison among those schemes. In the algorithm, three characters n, m and p are used as parameters. Character n means that in this scheme, data would be cut into n shares. Character m tells us under this scheme, m shares of original data are needed to reconstruct the data. Character p is the value with less than which the confidentiality of data would not been disclosed.

Table 1. Distribution Schemes

Algorithm	Parameters (n-m-p)
Non-Distribution	1-1-1
Replication	n-1-1
Data Striping	n-n-1
Splitting	n-n-n
Information Dispersal	n-m-1
Secret Sharing	n-m-m

a) Non-Distribution

It refers to traditional no-redundancy storage method. Data is solely stored in one server. Availability of data heavily relies on the availability of the server. Also, it is obvious that when the server is compromised, all information would be disclosed.

b) Replication

As the name implies, replication refers to the complete copy of the original data. It, to some extent, solves single point of failure (SPOF). Several replications of the original files guarantee data redundancy and effectively increase the availability of the original files. There are some inevitable and unpredictable reasons that may cause network inaccessibility or machine outage, which will result in data loss or render the data stored in cloud unavailable. Generally speaking, more replicas of data equals to higher availability of data. Besides, making replications for users' data can not only enhance the availability of data but also boost system performance by reasonably allocating replicas to storage nodes and realizing the nearest visit via certain configuration of router. Additionally, parallel visits can be realized when clients request for some data which is replicated and stored in several nodes. In this way, users can access data with less time-lag and the load of every node is relieved as well.

However, it is economically unfeasible to replicate all files as much as we want. Numerous replicas mean enormous consumption of system storage resources and

complicated file management. So a good balance must be struck among availability, system performance and cost. Another concern is that when the data is manipulated, all copies need to be modified. The last problem may be the increased risk of information disclosure. Because each copy contains the complete information of original data, single server compromising is fatal to confidentiality of entire data set. In cloud environment, replication is even useless, because customers would like the CPS to learn as least as their information and hold as least copies as possible.

c) Decimation (Striping)

RACS [8] identifies two kinds of failures in data storage: Outages and economic failures (vendor lock-in). In order to address the two problems, a redundant stripping mechanism is recommended. The technique adopted in redundant striping is called erasure code and the basic idea is that the data to be stored can be broken into N shares, each of which is 1/N size of the original data. By utilizing some coding techniques, the N shares are transformed to K blocks of data and then they are stored in K nodes. The total overhead factor of such storage is K/N. One only needs N available blocks to retrieve the original data.

The advantages of this approach are: It allows K-N blocks to be out of work; Provide higher fault tolerance rate and lower storage complexity. Disadvantage is that it requires complicated coding process and more computation. Compared to replication, the erasure coding is more likely to meet the needs of data storage which requires higher reliability and less storage cost.

d) Splitting

Data splitting is an approach to prevent critical information or confidential data against unauthorized users. The data is split into N shares and each share is encrypted and stored in different servers. To retrieve the data blocks from the servers, the user must know the locations of the nodes storing the parts. After all the split data is retrieved, the parts are combined together and decrypted. All these steps are completed on the basis that the user is an authorized one. All the n shares of data will be needed to reconstruct the original data. If one part is missing, the original data is gone forever.

e) Information Dispersal

Mo [9] proposed a well-known information distribution scheme - the Information Dispersal Algorithm (IDA). The IDA has been introduced in several cloud infrastructures. Bowers [10] proposed the High-Availability and Integrity Layer (HAIL), which can prove a user IDA distributed data's integrity in cloud. It uses the IDA to distribute files among several CPSs. A so-called "Dispersal Code" is used for checking data integrity.

Under this algorithm, information is cut into n pieces and distributed among n servers. Unlike data striping or splitting algorithm, IDA allows reconstruction of original information by a smaller m pieces than all n pieces. It means that even if several servers down, as long as at least m servers are alive, the original data can also be obtained. The benefit of IDA compared to data striping and splitting is obvious - It increases the availability of distributed data.

f) Shamir's Secret Sharing

Secret Sharing is a method created to solve secret information distribution among a group of parties. The Shamir's Secret Sharing scheme is widely used in information distribution area. In [11], a Multi-clouds Database Model based on the Shamir's Secret Sharing algorithm was proposed. Data would be cut into several pieces and be stored in several databases, which are under control by a unified DBMS layer.

Using this method, one can distribute data among n servers. Similar to IDA, Shamir's Secret Sharing allows reconstruction of original data by a subset, like m, of all n pieces. The difference is that, any single piece of Secret Sharing reveals no information about the original data. One piece of IDA algorithm contains a part of original data's information which is not complete but makes sense to the user. The advantage of Shamir's Secret Sharing algorithm is that a single point compromise reveals nothing about the data which provides more confidentiality than IDA does.

4.3.4 Measurement

Availability Assessment

Suppose that, the probability of a single node's uptime is r. We assume the availabilities of single nodes are independent. Then we can use the following formulas to calculate the probability of failure for each distribution algorithm [7].

Table 2. Availability Assessment

Algorithm	Availability
Non-Distribution	r
Replication	$1-(1-r)^n$
Data Striping	r^n
Splitting	r^n
Information Dispersal	$\sum_{i=m}^{n} \binom{n}{i} r^i (1-r)^{n-i}$
Secret Sharing	$\sum_{i=m}^{n} \binom{n}{i} r^i (1-r)^{n-i}$

Risk of Disclosure

Suppose the probability of single node compromise is p. We assume the disclosure risks of single nodes are independent. The risk of information disclosure for each algorithm can be calculated by the following formulas:

Table 3. Risk of Disclosure

Algorithm	Risk of disclosure
Non-Distribution	p
Replication	p × n
Data Striping	p × n
Splitting	p^n
Information Dispersal	p × n
Secret Sharing	p^m

We cannot tell which scheme has the largest availability or the smallest disclosure risk unless specific parameters, like n, m and p, are given.

4.3.5 Integrity Realization

Sanitization

Zhang & et al [12] put forward a new concept---cloud shredder, in their paper published in 2011. They emphasized the danger of losing critical information caused by the lost or theft of laptops, PCs, tablets or even hard drives. The idea specified is that every file in our devices (PCs, tablets, laptops, etc.) is split into two parts, one of which is stored in the local device while the other one is uploaded and stored on a remote server in the cloud. In this way, even though the devices are lost or stolen, the data in them will still remain safe because all the thieves get from the devices are pieces of useless data. If the other half of data in the cloud is hacked by bad people, they need the devices to view the complete version of data. As a result, the cloud shredder concept makes the data safer than storing all our secrets in the local devices or in the cloud.

However, several drawbacks are obvious with such a system. Backup is hard to realize; if either part of the data is lost (due to cloud failure or machine failure), how can we retrieve the complete data? If the internet is out of reach, how can we check the data? We cannot take ourselves with our laptops everywhere and anytime, so what if we want to access our data then? Is the cloud shredder better than a thumb drive?

The value of cloud shredder is that it provides a way to delete the data. If the laptop has been stolen, all we need to do is to delete the data in the cloud to make sure the thief can never obtain our secrets via our laptop.

Active bundles [13] can protect sensitive or private data from being disclosed to unauthorized or un-trusted parties and from being maliciously disseminated. The basic idea of active bundles is that an active bundle comprises of three parts: Sensitive data (the data we want to protect), metadata, and a virtual machine. The metadata is used to specify how the entire or parts of sensitive data can be accessed and how the active bundle can be disseminated. The virtual machine (or VM) uses information provided by metadata to manage the use of the active bundle. The three main operations of the active bundles are evaporation, apoptosis and self-integrity check. Evaporation means that after the active bundle arrives at a host and gets the trust level of it, the bundle will decide whether the sensitive data can be presented to the host and which part of this sensitive data can be released. The rest of the sensitive data which the host is not authorized to access is evaporated. Apoptosis means that if a bundle checks the potential danger of data compromise, it will perform self-destruction and leave no trace for malicious attackers. Self-integrity check means that the active bundle can use its own algorithm to check whether the integrity within its data is being compromised or not. Actually, the active bundle may be an effective solution to data protection in cloud due to its structure, mechanism, algorithm, encryption technology and so on. Also, it achieves the goal of not using un-encrypted data to do authentication and protecting identity information from un-trusted hosts as well. However, the whole concept of the active bundle solution is based on the assumption that the virtual machine is secure against attacks. In fact, the virtual machine can be a weak link or a loophole in the whole solution. So how to realize this assumption becomes the problem.

Single Node Independence

When we are applying distribution scheme in cloud, there is a question that should we distribute data either among one CPS's server or among several clouds. In [3], data is distributed among CPSs. In [11], data is distributed among servers.

Amazon S3 uses the concepts of Availability Zone (AZ) and Availability Region (AR) to illustrate its redundancy storage strategy [14]. ARs are geographically separated, like data centers in multiple locations. AZs are physical separated, like isolated servers in a data center. CPSs are more likely to be "business independent", which means they are isolated in ownership.

We use this concept to illustrate the differences between storage locations. Different method provides different single point independence. As what is discussed above, a comprehensive comparison including other business concerns, like cost-benefit analysis, would be needed.

4.4 Trust Mechanism in Cloud Storage

4.4.1 The Development Status and Trust Issues of Cloud Computing

In today's competitive environment, cloud computing offers a range of services which are so convenient that a large number of enterprises have been attracted by its highly scalable technology. Although enterprises gain many opportunities provided by cloud computing, new challenges cannot be ignored. Trust issue is one of the most obvious challenges which are a paramount concern for most enterprises. Actually, the dearth of customer confidence is not only comes from technology itself, also comes from a lack of transparency, a loss of control over data assets and unclear security assurances. The issues can only be resolved from two aspects, technology and psychology.

The most significant issue is the sufficiency of data information which can be presented with services. Other issues which are also very important, including control, ownership, prevention and security need to be considered. According to the research, we know that people not having confidence on cloud computing is not because they do not trust the service provider, but they feel their data is out of control. The ownership of data assets also has an impact on trust. When the enterprise consigns their data to the service provider, both the enterprise and its client should trust the cloud provider. If the client suspects the provider, once the data is damaged, the enterprise must take full responsibility. As for prevention, it is more important than compensation. In cloud storage, data security must be guaranteed. Although the corporation can compensate for the economic loss of the client, a security breach of data is irreparable. Obviously, no amount of money can remedy the lost data or the enterprise's reputation. Thus, more attention should be paid to prevent failure than to post-failure compensation.

Security, which plays a crucial role in preventing service failures and cultivating trust in cloud computing is urgent to be improved. Much effort has been done to satisfy customers' demand but the security status is lack of transparency. It means that the client is not clear on how service provider secures and controls their data, so they still hold a skeptical attitude towards data security.

There are two main challenges we confront today: Diminishing control; and Lack of transparency. Some data provided by clients need to be processed or stored on various disks in multiple locations and possibly managed by third-party providers. In this situation, because of the loss of control over the data and processes, data confidentiality, integrity and availability cannot be guaranteed. Additionally, in some cloud service model, the service provider usually has complete control of data, while enterprises retain only partial control of their data, which they often find quite alarming. Therefore, enterprises should be permitted to trace their data processing procedure. Actually, they do not need to really control or process their data. They only want to know where the data is and how it is dealt with which can improve their trust level of cloud computing.

Generally, because of the lack of transparency, the consumers perceive that an in-house system is more secure compared with cloud computing services. If we improve the level of transparency, we can decrease consumes' perception. There are two issues existing in transparency: One is the physical location of storage and processing sites; the other is the security profiles of these sites. Therefore, in order to address the issue, the service provider must supply sufficient information of their data through which the client can trace the location and processing procedure of the data.

4.4.2 A Model Applied to Build a Trust Environment in Cloud Storage

Trust is a subjective measurable scale that can thrust decisions based on the beliefs of decisions [1] and it is widely used in social science to build human being's relationship. However, trust should not be restricted within the domain of philosophy, sociology and psychology. It needs to be addressed by all attempting good governance [2]. Nowadays, trust is becoming an essential part to form a secure distributed cloud computing environment, especially in cloud storage, because trust has many security properties, such as reliability, dependability, confidence, honesty etc. In cloud environment, trust issues increase because customers feel their data is not under control. As a customer's infrastructure is located at an off-site location and managed by a second or third party entity, the customer lacks the confidence of transparency, data security and unclear security assurance. Therefore, we can regard trust in cloud as the customers' level of confidence in using the cloud. In order to enhance customers' confidence level, we need to increase their competence trust and benevolence trust of cloud computing service by mitigating technical barriers and improving security management.

The above provides us a concept that how we can construct a trust mechanism in cloud computing to encourage people to use cloud computing service. "Trust decision usually consist of two parts, reasoning and feeling, both cognitive trust and emotional trust will influence the decision making process" [7]. In the emotional part of trust, it contains more about individual's feelings and emotion. [15] pointed out that "when people feel that the trustee is well-meaning and has the intention to work for trustor's benefit, he/she will think the trustee is benevolent" [15]. Such beliefs of benevolence is necessary for an emotional trustworthy relationship which make the trustor feel that he/she will not be cheated and have less risk, in turn enhance trustor's feelings of

comfortable and secure. Thus, we propose that the benevolence trust has a positive influence on emotional trust.

[16] also indicated that "customer trust in trustee's competence means that the customer believe that the trustee has the capability to transmit professional and helpful information, which also mean that the customer will be more relying on the relationship between them for decision make and feel comfortable and secure" [16]. Thus, customer's cognitive trust in trustee's capability will influence their emotional trust in trustee, according to that; we propose that the competence trust has a positive influence on emotional trust. The level of customers' emotion trust will influence their consuming behavior.

In order to improve people's trust level of cloud computing service, we need to improve both benevolence trust and competence trust. Benevolence trust reflects a person's psychology statement and the lack of it comes from the uncertainty between two parties. If no uncertainty exists between two parties, it indicates that no risk or threat is found in future interaction between two parties [3]. However, we live in a real world in which we cannot absolutely eliminate the uncertainty and what we can do is to reduce uncertainty and to increase predictability on what other party will act in the future. The way to decrease uncertainty is the communication and the share of information between parties [4]. According to Berger [17], uncertainty about the other party is the "(in)ability to predict and explain actions". It means that uncertainty can be reduced by sharing information and obtaining the condition of the information [17]. The reason why customers do not trust the service provider is that they know nothing about their data processed by the provider. When the valuable data is put in total dependence of someone else, the customer feels unsafe and the level of uncertainty increases. Therefore, the provider of cloud storage needs to improve the level of transparency, which means they should let customers know what will be done on their data and keep customers tracking the situation of the data.

As for competence trust, it is a cognitive feeling to measure someone's ability in specific area. In cloud storage, if the service provider wants to improve customers' competence trust for them, they must enhance their secure technology to protect users' data from three aspects which are confidentiality, integrity and availability.

5 Conclusion

Cloud computing is considered to be one utility service after water, gas and electricity power. As an important component of cloud computing, cloud storage is facing various security challenges. To protect the valuable data stored in cloud, we proposed several possible solutions to cloud storage security problems in terms of confidentiality, availability and integrity.

Encryption as a traditional method for data confidentiality protection has been proved very effective in cloud environment. Although, problems remains, such as manipulating encrypted data in cloud cause computational overhead. Several intelligent solutions to those problems have been proposed. Encrypting data before it is sent to the storage pool is the most suggested way used to solve untrusted CPS problem.

Other than confidentiality, the availability of data in cloud also attracts attention. As a utility service, cloud should be able to provide reliable service at any given time. To meet with their own needs, users can use those straightforward comparisons combined with non-technical concerns to make a better decision.

One important concern on data integrity in cloud is how users can know their data is integrated in cloud. In the past, people did not pay more attention to build a perfect trust mechanism and more concerns are concentrated on developing technology. Recently, we found that various techniques have been created to protect the data security, but no matter how advanced the technology is, the customers still suspect the security of the data. Therefore, we must use the knowledge of psychology and information technology to build the trust mechanism. Integrity of data can be delivered by both non-technical and technical methods. A well-designed trust mechanism provides users the basic integrity protection in cloud. In cloud environment, key management is related to different parties. To enhance the trust between those parties, a third party scheme of key management is proper.

References

[1] Brohi, S.N., Bamiah, M.A., Brohi, M.N., Kamran, R.: Identifying and analyzing security threats to Virtualized Cloud Computing Infrastructures. In: 2012 International Conference on Cloud Computing Technologies, Applications and Management (ICCCTAM), December 8-10, pp. 151–155 (2012) ISBN: 978-1-4673-4415-9

[2] CIOTIMES. 云计算安全架构中的加密. TechTarget China (October 19, 2012), http://www.ciotimes.com/cloud/cjs/72702.html

[3] Han, S., Xing, J.: Ensuring data storage security through a novel third party auditor scheme in cloud computing. In: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), September 15-17, pp. 264–268 (2011) ISBN: 978-1-61284-203-5

[4] Brodkin J. : Gartner: Seven cloud-computing security risks. NetworkWorld (July 2, 2008), http://www.networkworld.com/article/2281535/data-center/gartner-seven-cloud-computing-security-risks.html

[5] Whitman, M.M., Mattord, H.J.: Principles of information security, 4th edn. Cengage Learning (January 1, 2011) ISBN: 978-1111138219

[6] Singh, Y., Kandah, F., Zhang, W.: A secured cost-effective multi-cloud storage in cloud computing. In: 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), April 10-15, pp. 619–624 (2011) ISBN: 978-1-4577-0249-5

[7] Wylie, J.J., Bakkaloglu, M., Pandurangan, V., Bigrigg, M.W., Oguz, S., Tew, K., Williams, C., Ganger, G.R., Khosla, P.K.: Selecting the Right Data Distribution Scheme for a Survivable Storage System. Carnegie Mellon University (May 2001), http://www.pdl.cmu.edu/ftp/Storage/CMU-CS-01-120.pdf

[8] Abu-Libdeh, H., Princehouse, L., Weatherspoon, H.: RACS: a case for cloud storage diversity. In: SoCC 2010 Proceedings of the 1st ACM Symposium on Cloud Computing, pp. 229–240 (2010) ISBN: 978-1-4503-0036-0

[9] Rabin, M.O.: Efficient dispersal of information for security, load balancing, and fault tolerance. Journal of the ACM (JACM) 36(2), 335–348 (1989), doi:10.1145/62044.62050

[10] Bowers, K.D., Juels, A., Oprea, A.: HAIL: a high-availability and integrity layer for cloud storage. In: Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 187–198(2009) ISBN: 978-1-60558-894-0

[11] AlZain, M.A., Soh, B., Pardede, E.: MCDB: Using Multi-clouds to Ensure Security in Cloud Computing. In: 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing (DASC), pp. 784–791 (2011) ISBN: 978-1-4673-0006-3

[12] Zhang, N., Jing, J., Liu, P.: CLOUD SHREDDER: Removing the laptop on-road data disclosure threat in the cloud computing era. In: 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), November 16-18, pp. 1592–1599 (2011) ISBN: 978-1-4577-2135-9

[13] Ranchal, R., Bhargava, B., Othmane, L.B., Lilien, L., Kim, A., Kang, M., Linderman, M.: Protection of identity information in cloud computing without trusted third party. In: 2010 29th IEEE Symposium on Reliable Distributed Systems, October 31- November 3, pp. 368–372 (2010) ISBN: 978-0-7695-4250-8

[14] Amazon EC2. Availability Region and Availability Zone concept, `http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/using-regions-availability-zones.html`

[15] McKnight, D.H., Choudhury, V., Kacmar, C.: The impact of initial consumer trust on intentions to transact with a web site: a trust building model. Journal of Strategic Information Systems 11, 297–323 (2002)

[16] Komiak, S.Y.X., Benbasat, I.: The effects of personalization and familiarity on trust and adoption of recommendation agents. MIS Quarterly 30(4), 941–960 (2006)

[17] Dean, J., Ghemawat, S.: MapReduce: Simplified Data Processing on Large Clusters. Communications of the ACM 51(1), 107–113 (2008), doi:10.1145/1327452.1327492