Cloud Computing: Data Storage Security Analysis and its Challenges

Navdeep Aggarwal, Parshant Tyagi, Bhanu P. Dubey and Emmanuel S. Pilli
Department of Computer Science & Engineering
Graphic Era University, Dehradun, India

ABSTRACT

Cloud computing is emerging as a new technology in the field of computing which effects many aspects of computing. The users are enabled to move their data and application software to the network and access the services on-demand. Moving data outside the organizational boundaries and access them through internet makes complete loss of control from the owner's side. The data over the cloud servers is easily accessed by both the computing providers and law enforcement agencies easier in comparison to data stored on your local disk. Cloud brings many different security issues and among them data security acts as one of the major challenge in cloud computing. Data security becomes an important issue for securing outsourced data and to maintain a level of trust among data owners. In this paper we will analyze the security requirements and the various approaches for data security. We also highlight the new emerging research challenges in data security and privacy.

Keywords

Cloud computing, data privacy, cloud data storage, security

1. INTRODUCTION

Cloud computing realizes computing as a utility. It provides a pool of resources which can be allocated to users dynamically according to their requirement. Thus both the users and providers are benefited: providers can reuse their resources and users acquire and release resources according to their requirement [1].

The NIST working definition summarizes cloud computing as: "*a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction*". According to the definition the cloud computing has the following essential characteristics. The cloud provide on demand self-service in which user can provision the resource (network, storage, computing) whenever required without human interaction. Computing facilities are available over the internet which can be easily accessed by the devices like mobile phones, laptops, PDAs anywhere and at any time [2]. Service provider in cloud computing enables a pool of resources which include the physical as well as virtual devices according to the user requirement and provide multi tenant environment for better utilization of underlying hardware. Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. In cloud computing business firms, end-users pay for what they have used. Pay as you go plan is implemented. A parallel to this concept can be drawn from the electricity and water distribution where user is not aware of the underlying infrastructure and devices required to provide the service.

In practice, cloud service can be grouped into three categories: software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). SaaS refers to the software available on the internet. Cloud providers enable application software as on-demand services. In this the end user has no need to download, install, configure and maintain the software. Examples of SaaS providers include google applications, Salesforce.com, youtube, facebook. PaaS refers to the platform and enables programming environment in which developers customize application. Examples of PaaS providers include Amazon DB/S3[3], Microsoft Azure, Google AppEngine. IaaS refers to the set of virtualized infrastructure components such as virtual machines, network bandwidth, storage and necessary tools to build an application environment on which customers build and run applications. Examples of IaaS providers include Amazon EC2, GoGrid, FlexiScale.[4]

The reminder of this paper is organized as follows. In section 2, we describe the background. Data security requirements are examined and a detailed survey of the current technologies used for cloud computing is done in section 3. In section 4, various data security approaches are analyzed. Current research challenges of data security are described in section. 5. Finally the paper concludes in section. 6.

2. BACKGROUND

Cloud computing related technologies include grid computing, utility computing, virtualization and autonomic computing. Cloud computing is similar to Grid computing in a aspect that here also resources are coordinated to achieve common computational objective but it is one step ahead that it leverage virtualization technology for better resource utilization and dynamic resource provisioning. Cloud computing acts as a realization of utility computing which include on-demand resource provisioning and utility based pricing scheme. Cloud computing relates to autonomic computing in a way that it support autonomic resource provisioning but its objective it to reduce cost neither to reduce system complexity [5]. Cloud computing shares characteristics of other computing technologies hence present unique benefits over other technologies but at the same time new security issues arises.

Data security is the major challenge in the cloud computing since the customer's data and business logics reside in the servers which are remotely situated and far away from the end-users. These data may include confidential data (financial data, health records), personal information which may be disclosed to competitors or publicly. So security emerges as the highest priority issue [6].

According to a survey from IDCI in 2009, 74% IT managers and CIOs believed that the primary challenge that hinders them from using cloud computing services is cloud computing security. Another survey carried out by Garter in 2009, more than 70% CTOs believed that the primary reason not to use cloud computing services is that there are data security and privacy concerns[6]. As a result security issues act as a major hindrance in the implementation of cloud computing. The security requirements fall into four categories confidentiality, integrity, recoverability and availability. Data confidentiality is the major requirement for which the end-users worry.

Security is one of the most difficult task to implement in cloud computing. CISCO CEO John Chambers said "Cloud computing is a security nightmare and it can't be handled in traditional ways". Cloud data storage is build of thousands of cloud storage devices clusters by network, distributed file system and other storage middleware to provide storage [7]. CDS however offer services to assure integrity of data. But don't provide solution to data integrity problem. So the users itself have to adopt some data security technique to assure data integrity in cloud computing.

3. DATA SECURITY REQUIREMENTS

3.1 Data Integrity

Data integrity may be easily preserved by employing traditional cryptographic methods such as message authentication codes (MAC). It is a fixed size block of data based on file F using any secret key. The data owners maintain small amount of MAC before outsourcing the data and whenever data is needed, MAC is verified with the previously computed MAC to verify the correctness of the received data from cloud.

In cloud environments traditional methods are implemented no more since the data is dynamic and there is huge cloud storage [8]. So it becomes quite impractical that for checking whether the data is stored securely we retrieve all the data stored on server. Integrity threats include data deletion, data manipulation. The computation details are not transparent to cloud customers so the CSP behave dishonestly and may alter the data.

3.2 Data Confidentiality

A user can access the services provided by SaaS model through web browser over the internet. So to protect the data during transmission HTTPS is implemented. If user uploads data to the CSP security must be there so that only authorized users access the data, the same requirement with PaaS. While in IaaS, multiple users data reside on the same location so in IaaS confidentiality arises in a way to include isolation over the different user's data.

3.3 Data Availability

Availability is affected if the server or service organization is penetrated or spoofed. In cloud characteristic broad network access DNS is one of the main attack on availability [1]. So for better service offering over the internet user must have reliable DNS. So for long term data storage data availability is of much importance due to possibility of data loss.

3.4 Data locality and Recovery

In cloud environment data is dynamically stored over the cloud servers, so the user is unaware where its data is going to be stored. Due to privacy concerns some EU countries forbid the storage of sensitive data outside the country boundaries. So the data locality is an important issue [9]. Another question arises if some investigation occurs, under whose jurisdiction the investigation occurs. The issue can be solved by creating a secure SaaS model which provides reliability of the location the data.

The data in cloud may become unavailable or effected due to environmental disaster, server failure. For this recovery of important documents must be maintained either by user on its local disk or backup services are provided by many cloud vendors like Amazon S3.

3.5 Data Breach

Due to Multi tenancy environment in cloud breaching the data will become a potential threat. Data breach effects two security properties of data confidentiality and integrity. Confidentiality refers that only authorized parties or systems can access the data and integrity refers that data is not deleted, manipulated or fabricated by some third party who is not authenticated to perform such task. Data breach may occur internally by some data manager who has direct access to the data or from outside by some malicious hacker. However confidentiality and integrity issues are addressed by strong cryptographic mechanism like DES and AES with common PKI infrastructure. In this data and key management become an issue for data owner which can be addressed by combining techniques of attribute based encryption, proxy re-encryption and lazy re encryption.

3.6 Data Segregation

Multi tenancy is an important characteristic of cloud computing. In multitenant, multiples users and organizations reside at the same location. So it becomes possible for the malicious users to gain access to the other users data. So keeping data separate and maintain isolation among the users is an important issue [10]. These issues can be solved either by creating a robust virtualization platform or by implementing Trusted Platform Module embedded on the motherboard.

4. STATE OF ART

This section provides a review of approaches used in cloud computing environments. Various cryptographic protocols are implemented to preserve privacy in cloud computing and to provide data security in service delivery models. Robust virtualization environment is created through various trusted computing technology for VM security in IaaS delivery model.

4.1 Trusted Computing Approach

Jinzhu et al. [11] presents a practical architecture to protect data confidentiality for guest virtual machines by constructing virtualization platform. This is built by hypervisor Xen and Trusted platform module (TPM) integrated on motherboards. Before boot the guest virtual machine, user prepare encrypted disk image of root file system. Then prepare boot disk image, install grub on disk and put kernel on it. Send these two disk images to the dom0 on the cloud server.

Szefer et al. [12] propose a cyber-physical security frameworks for data centers that combine security mechanisms in cyber and physical space. Defense framework is based on the time difference between the attack detection and the actual attack. Based on the physical devices like sensors it is possible to initiate the defense mechanisms which

include delete, encrypt and move. The defense mechanism is adopted according to the requirement of the end-user and according to the type of information stored on the server.

Siqin et al. [13] propose logging VM for secure logging of auditable File System. Auditing logs are isolated in another VM on the same host. By the isolation provided by VMs, the logs are kept safe in another VM even if the working VM get crashed. So the privileged user can't access, modify or delete data of unprivileged user.

Fu et al. [14] analyses cloud based virtualization concerns which include VM security and threats, Hypervisor Security, Data Leakage, Privacy, Data remanence issue in virtualization and attacks in virtualization level.

Jasti et al. [15] analyses the security threats that compromise the virtual machine and the hypervisor itself. These threats include VM hoping, VM escape and mobility. In multi-tenancy environment data of multiple organizations and user reside on the same physical location. In such systems resources are transparently shared among VM of different users. So a malicious user having control over VM try to gain access over other VM or try to compromise with data of other VM. For this usage of resources such as C.P.U, memory and network are analyzed on VM by creating a virtualized environment.

4.2 Information Centric Security

Xiao et al. [16] presents frame work to ensure data security in cloud storage system. SLA is used as the common standard between user and provider and several technologies are discussed to make data stored in cloud safe. These technologies include storage protect, transfer protect and authorize. For secure storage data is divided to small pieces and save them to different places. If data pieces in one data center crashed, the data can be resumed by left pieces. Cryptographic protocols like SSL and TLS provide security for communication over networks such as internet.

Deyan et al. [6] Data security and privacy protection issues in cloud around data life cycle that is from generation to destruction of data. Like in storage phase data stored in cloud need three security requirements as confidentiality, integrity and availability. The confidentiality issue is solved by encryption algorithm and key strength. Similarly due to physical characteristic of storage medium data deleted is still exists and can be retrieved.

Xiaojun et al. [17] describe data security through data life cycle process. The process includes stages that are creating, store, use, share, archived and destruct. Firstly client proxy generates data and classifies it by marking. For storing data to the cloud, user either encrypts it before transferring or sends it through secure network. After receiving data server checks its integrity. If client want to use and share data integrity proof is given by the server. If the data is not used for longer time then achieve request is send to the server and it may be transferred to another location and if it is not to use anymore it is destroyed.

Zhifeng et al. [18] proposed Auditable MapReduce for the current MapReduce model for making cloud platform trustful. In map reduce scenario, all the machines (mappers and reducers) are responsible for performing task. MapFunction takes key pair input and generate output (intermediate key pair). These intermediate results are sorted by key and taken as input by reduce function and generate final output. A-test are performed and compare output from worker and from auditor to determine whether the worker is malicious or not. Due to high overhead in A-test it provides P-Accountability since with lower number of malicious nodes less number of records to be checked.

4.3 Cryptographic Protocols Approach

Sood [19] proposes a security model for the whole computing process in the cloud. The characterization and measures are presented for secure storage and efficient retrieval of data from cloud. Data is classified based on cryptographic parameters like confidentiality, availability and integrity. Based on Sensitivity rating data is distributed in public, private and limited access section. Data and index are encrypted with the 128 bit SSL encryption. MAC is generated to check whether the data is not tampered for that calculated MAC is matched with the MAC received along with the encrypted file. The model is analyzed and tested with the help of cloud computing simulator Hadoop. The proposed framework is compared with the various security issues that hinders the cloud computing from adopting it.

Jian et al. [20] address the issues of privacy problems in cloud computing like disclosure of sensitive information which include personal identification information, unauthorized access to personal data, protect privacy of data when moving outside the organization boundaries. Anonymity algorithm is proposed in which data is processed by this algorithm before it is send to service provider. SP integrate auxiliary information to analyze data. The anonymity method is different from cryptographic technology here the service provider directly use the data without any key.

Xu et al. [21] analyses the architecture and layers of cloud computing. Analyses the security of cloud computing platform on which applications are deployed. Design security framework for the cloud computing platform on security issues like confidentiality, integrity, availability, non-repudiation and reliability. The operation process and modules description of security framework is done.

4.4 Identity Management Approach

Wang et al. [23] proposes private matching protocol for matching the personal information to the tuples of the data sets in the data center. The user checks whether the anonymized data added to the table sets meet K-anonymity. If it meet the K-anonymity data generalization is done through minimal attribute generalization approach to overcome information loss while satisfying service provider request and enables it to send the service to right client.

Ranchal et al. [24] proposes Identity Management Approach (IDM) for data privacy and security which is independent of Trusted third Party. The approach makes use of predicates over encrypted data and negotiation for using cloud service. For privacy active bundle scheme is implemented which acts as a middleware agent that include PII data, privacy policies and set of protection mechanisms. It allows use of IDM application over untrusted hosts.

5. RESEARCH CHALLENGES

- The data is dynamically stored over the cloud servers. Since the data in cloud is not only confined to organization boundary but also stored outside the organizational boundaries which gave raise to many security challenges. The challenge is if investigation occurs under whose jurisdiction the investigation occurs.

- For secure data access and to provide confidentiality various encryption techniques are implemented. The major challenge that came across is scalable key management.To overcome security threat a trustable entity must be there. This entity may be a third party auditor that has expertise and capabilities to access cloud storage security on behalf of data owner. So challenge is to appoint a third party which is trusted by both the data owner and cloud service provider.
- The biggest challenge is to keep a view that service provider follows the terms and policies in providing services to the customers as per defined in the Service Level Agreement (SLA) established through negotiation
- between cloud service provider and user. All the data security laws are followed in order to provide a secure and robust environment.
- Privacy and identity management is another necessary concern as customers data reside on cloud distributed servers which are owned and managed by cloud service providers. So protecting personal information and confidentially data (financial data, health record) from being disclosed to public or business competitors is a big challenge.

Table 1. Data Security Approaches

No.	*Approach*	Description	Example
1.	*Trusted Computing Approach*	End users are trusted that the data residing in the VM and the configuration files are secure, attacks are analyzed with N/W, memory and C.P.U usage.	Hierarchical secure virtual model[22], virtualization approaches[14], robust virtualization platform [11].
2.	*Information Centric Security*	Security is provided from data generation to data destroys. SLA are managed and verified that the services are delivered according to the terms and conditions defined in SLA.	Data Security framework [16], cloud computing security architecture [6]
3.	*Cryptography and Key management*	Data is classified based on cryptographic parameters CIA, Sensitivity rating is calculated and encryption algorithm is applied.	Security evaluation through encryption algorithm [19], crypto cloud storage scenario [9]
4.	*Identity Management Approach*	The system likely control what PII it revel about itself, and who can access that information.	Private Matching Protocol[23] ,Active Bundle Scheme[24],

6. CONCLUSION AND FUTURE WORK

The data security is the major challenge in cloud computing which hinders the organizations and industries in acquiring cloud services. The security vulnerabilities always create a fear of data loss and leakage. The data security issues exits in all the service delivery models. Though the cloud computing realizes many advantages which include cost- efficient, flexibility, requires very less initial setup for starting organizations

To address security and privacy issues in cloud computing, we understand the vulnerabilities of our system and try to implement certain framework on the cloud computing services.

7. REFERENCES

[1] T. Hsin-Yi, M. Siebenhaar, A. Miede, H. Yu-Lun, and R. Steinmetz, "Threat as a Service?: Virtualization's Impact on Cloud Security," IT Professional, vol. 14, no. 1, pp. 32-37.

[2] Z. Xiao and Y. Xiao, "Security and Privacy in Cloud Computing," Communications Surveys & Tutorials, IEEE, vol. PP, no. 99, pp. 1-17.

[3] Amazon.com, (2008), "Amazon Web Services(AWS)"," Available: Online at http://aws.amazon.com,

[4] S. Sakr, A. Liu, D. M. Batista, and M. Alomari, "A Survey of Large Scale Data Management Approaches in Cloud Environments," IEEE Communications Surveys & Tutorials, vol. 13, no. 3, pp. 311-336.

[5] Q. Zhang, L. Cheng, and R. Boutaba, "Cloud computing: state-of-the-art and research challenges," Journal of Internet Services and Applications, vol. 1, no. 1, pp. 7-18.

[6] C. Deyan and Z. Hong, "Data Security and Privacy Protection Issues in Cloud Computing," in International Conference on Computer Science and Electronics Engineering (ICCSEE), 2012 pp. 647-651.

[7] A. M. Talib, R. Atan, R. Abdullah, and M. Azrifah, "CloudZone: Towards an integrity layer of cloud data storage based on multi agent system architecture," in Open Systems (ICOS), pp. 127-132.

[8] W. Cong, R. Kui, L. Wenjing, and L. Jin, "Toward publicly auditable secure cloud data storage services," IEEE Network, vol. 24, no. 4, pp. 19-24.

[9] P. You, P. Yuxing, W. Liu, and S. Xue, "Security Issues and Solutions in Cloud Computing," in Distributed Computing Systems Workshops (ICDCSW), , pp. 573-577.

[10] S. Subashini and V. Kavitha, "Review: A survey on security issues in service delivery models of cloud computing," J. Netw. Comput. Appl., vol. 34, no. 1, pp. 1-11.

[11] K. Jinzhu, "A Practical Approach to Improve the Data Privacy of Virtual Machines," in Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on, 2010, pp. 936-941.

[12] J. Szefer, P. Jamkhedkar, C. Yu-Yuan, and R. B. Lee, "Physical attack protection with human-secure virtualization in data centers," in Dependable Systems and Networks Workshops (DSN-W), 2012 IEEE/IFIP 42nd International Conference on, pp. 1-6.

[13] Z. Siqin, C. Kang, and Z. Weimin, "Secure Logging for Auditable File System Using Separate Virtual Machines," in Parallel and Distributed Processing with Applications, 2009 IEEE International Symposium on, 2009, pp. 153-160.

[14] W. Fu and X. Li, "The study on data security in Cloud Computing based on Virtualization," in International Symposium on IT in Medicine and Education (ITME), 2011 pp. 257-261.

[15] A. Jasti, P. Shah, R. Nagaraj, and R. Pendse, "Security in multi-tenancy cloud," in IEEE International Carnahan Conference on Security Technology (ICCST), 2010 pp. 35-41.

[16] Z. Xiao, D. Hong-tao, C. Jian-quan, L. Yi, and Z. Lei-jie, "Ensure Data Security in Cloud Storage," in Network Computing and Information Security (NCIS), 2011 International Conference on, pp. 284-287.

[17] Y. Xiaojun and W. Qiaoyan, "A View about Cloud Data Security from Data Life Cycle," in Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on, pp. 1-4.

[18] X. Zhifeng and X. Yang, "Accountable MapReduce in cloud computing," in Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on, pp. 1082-1087.

[19] K. S. Sandeep, "A combined approach to ensure data security in cloud computing," J. Netw. Comput. Appl., vol. 35, no. 6, pp. 1831-1838.

[20] W. Jian, Z. Yan, J. Shuo, and L. Jiajin, "Providing privacy preserving in cloud computing," in International Conference on Test and Measurement, (ICTM '09) 2009, pp. 213-216.

[21] X. Xu and J. Yan, "Research on Cloud Computing Security Platform," in Computational and Information Sciences (ICCIS), 2012 Fourth International Conference on, pp. 799-802.

[22] S. Manavi, S. Mohammadalian, N. I. Udzir, and A. Abdullah, "Hierarchical secure virtualization model for cloud," in Cyber Security, Cyber Warfare and Digital Forensic (CyberSec), 2012 International Conference on, pp. 219-224.

[23] Y. Z. Jian Wang, Jiajin Le, "A New Privacy Preserving Approach used in Cloud Computing," vol. Vols. 439-440, pp. 439-440, 2010.

[24] R. Ranchal, B. Bhargava, L. B. Othmane, L. Lilien, K. Anya, K. Myong, and M. Linderman, "Protection of Identity Information in Cloud Computing without Trusted Third Party," in Reliable Distributed Systems, 2010 29th IEEE Symposium on, pp. 368-372.

.