Survey on Various Techniques for Data Storage Security in Cloud Computing

Jahnvi S. Kapadia

Abstract—Cloud computing is a model for enabling convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. W ith its ability to provide users dynamically scalable, shared resources over the Internet and avoid large upfront fixed costs, cloud computing h as recently emerged as a promising hosting platform that performs an intelligent usage of a collection of services, applications, information and infrastructure comprised of pools of computer, network, information and storage resources. However along with these advantage s, storing a large amount of data including critical information on the cloud motivates highly skilled hackers thus creating a need for the security to be considered as one of the top issues while considering Cloud Computing. This survey paper aims to analyze various data storage security methods.

Index Terms— Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), Infrastructure-as-a-Service (IaaS), Security, Data Encryption, Compression, Authentication, Data Decryption, Decompression.

1 INTRODUCTION

Cloud computing overlaps some of the concepts of distributed, grid and utility computing, however it does have its own meaning if contextually used correctly. Cloud computing really is accessing resources and services needed to perform functions with dynamically changing needs. An application or service developer requests access from the cloud rather than a specific endpoint or named resource. What goes on in the cloud manages multiple infrastructures across multiple organizations and consists of one or more frameworks overlaid on top of the infrastructures tying them together. The cloud is a virtualization of resources that maintains and manages itself. Since the security is not provided in cloud, many companies adopt their unique security structure [10] .For e.g. Amazon has its own security structure. Here a detailed survey on different techniques for cloud security using encryption, decryption, compression etc has been done.

2 CLOUD TAXONOMY AND CHACTERISTICS

Cloud computing can be classified based on the services offered and deployment models. According to the different types of services offered, there are three major service models currently associated with cloud computing: Cloud Infrastructure as a Service (IaaS), Cloud Software as a Service (SaaS), Cloud Platform as a Service (PaaS).
The following summarizes the key concepts of each of these three service models.

2.1 Cloud Infrastructure as a Service (IaaS)

A model in which an organization outsources the equipment used to support operations including storage, hardware, virtual servers, databases, and networking components. The service provider owns the equipment and is responsible for housing, running, and maintaining it. The client typically pays on a per-use basis. It is a well known fact that except for a few peak times per year, most servers are running with a 7-10% load. IaaS enables enterprises to leverage the cloud during peak need times. Doing this is often referred to as cloud bursting. To accomplish this internally, organizations must use complex resource allocation software [1].

2.2 Cloud Software as a Service (SaaS)

The capability offered to the consumer is to use the provider's commercially available applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a Web browser. So applications remain available to end users when needed via the Web. One of the most common uses for SaaS is for Web-based email services. In addition, small to mid-size enterprises typical use SaaS for hosting commercial software suites such as customer relationship management (CRM), enterpriser planning (ERP), and supply chain management (SCM). SaaS enables enterprises to obtain the use of such commercially available software on demand without the need to invest in IT resources knowledgeable in its support [1].

2.3 Cloud Platform as a Service (PaaS)

The two components of PaaS are the place on which software can be launched (platform), and the services being provided (solution stack). Resources being delivered via PaaS typically include infrastructure and applications. In many cases the data being used is also stored in the cloud and the end user's terminal may contain only an operating system and Web browser. In addition, end users can write their own code and the PaaS provider then uploads that code and presents it on the Web. SalesForce.com's Force.com is an example. The PaaS model enables resources to be increased easily with demand since end users share the same cloud. This is often called multi-tenant cloud computing [1].

The characteristics often associated with cloud computing is as follows:

On-demand self-service by consumers: The technology is available to end users as it is needed without intervention by internal IT.

Broad access via the network: The technology can be accessed via an enterprise or public network and can be easily accessed from any location.

Resource pooling of physical and virtual resources: Since the data center and virtual resources are situated in one place, the technology can be shared and maintained more efficiently.

Rapid scaling of capacity: The cloud permits the rapid provisioning of end users and shared resources allows for elasticity of demand and capacity.

Enhanced transparency of usage via metrics: Centralized resources, vendor management, and contractual Service Level Agreements (SLAs) facilitate increased transparency and meaningful information that can be used by management.

3 LITERATURE SURVEY

3.1 Cloud Computing: Storage As A Service

Among other advances, cloud computing has brought advantages in the form of online storage. In this section, Storage-as-a-Service is referred. The range of service offerings in this space is remarkable, and they are continuing to grow. Data security for such a cloud service encompasses several aspects including secure channels, access controls, and encryption. And, when we consider the security of data in a cloud, we must consider the security triad: confidentiality, integrity, and availability. In the cloud storage model, data is stored on multiple virtualized servers. Physically the resources will span multiple servers and can even span storage sites. Among the additional benefits of such generally low-cost services are the storage maintenance tasks (such as backup, replication, and disaster recovery), which the CSP performs. The most notable provider in this space is Amazon with its S3 (Simple Storage Service). Amazon launched S3 in March of 2006. A common aspect of many cloud-based storage offerings is the reliability and availability of the service. Figure 1 depicts an abstracted view of how many individual disks in many aggregated storage devices are composed into a virtualized unit of storage. [6]

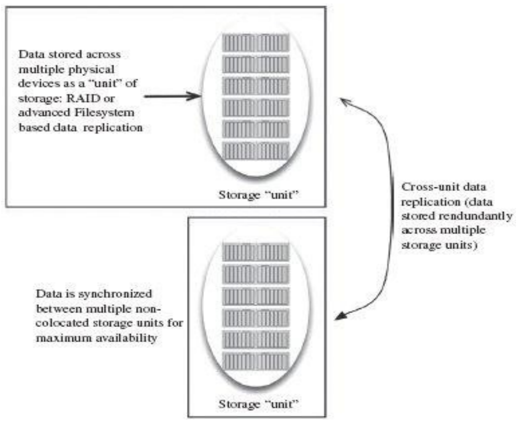
Data stored across multiple physical devices as a "unit" of storage: RAID or advanced Filesystem based data replication
Storage "unit"
Cross-unit data replication (data stored rendundantly across multiple storage units)
Data is synchronized between multiple non-colocated storage units for maximum availability
Storage "unit"

Fig 1: Cloud storage-replication and availability

Replication of data is performed at a low level by such mechanisms as RAID or by a file system. One such file system is ZFS, which was designed by Sun Microsystems as both a file system and a volume manager. ZFS supports high storage capacities and performs numerous security relevant functions including copy-on-write cloning and continuous integrity checking along with automatic repair. One of the more recent trends in online cloud-based storage is the cloud storage gateway. Several vendors offer such solutions that are generally implemented as an appliance that resides onsite at the customer premises. These appliances can provide multiple features, including:

Translation of client-used APIs and protocols (such as REST or SOAP) to those that are used by cloud-based storage services (such as NFS, iSCSI, or Fibre Channel). The goal is to enable integration with existing applications over standard network protocols.

Backup and recovery capabilities that work with in-cloud storage.

Onsite encryption of data that keeps keys local to the onsite appliance.

The vendors and products in this space include Gladnet, Nasuni Cloud Storage Gateway, StorSimple, and Emulex. The product and solutions that are available are seeing rapid changes and new functionality. Figure 2 depicts a typical cloud storage gateway application as it is used to augment local storage by acting as an onsite secondary copy and as an intermediary to the CSP storage service.

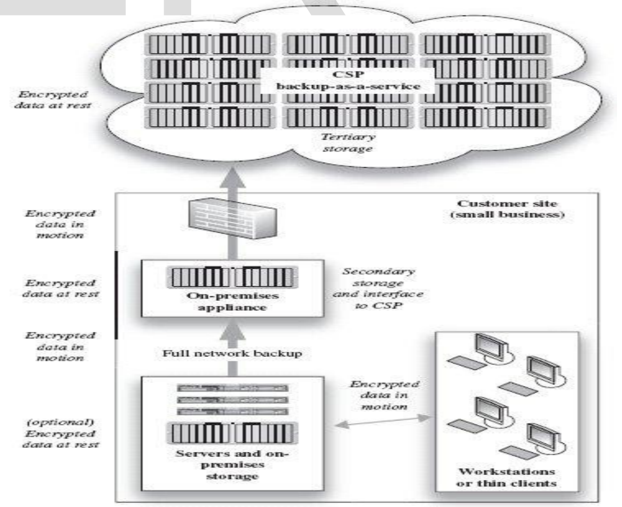

Fig 2: Cloud storage gateway appliance

3.2 Need for Cloud Security Techniques

Traditional cryptographic primitives for the purpose of data security protection can not be directly adopted due to the users' loss control of data under Cloud Computing. Therefore, verification of correct data storage in the cloud must be conducted without explicit knowledge of the whole data.

Cloud Computing is not just a third party data warehouse.

The data stored in the cloud may be frequently updated by the users, including insertion, deletion, modification, appending, reordering, etc. To ensure storage correctness under dynamic data update is very important.
The deployment of Cloud Computing is powered by data centers running in a simultaneous, cooperated and distributed manner. Individual user's data is redundantly stored in multiple physical locations which increase the data integrity threats.

3.3 Enabling Public Auditability And Data Dynamics For Storage Security In Cloud Computing

The third party auditor (TPA), who has expertise and capabilities that cloud users do not have and is trusted to assess the cloud storage service security on behalf of the user upon request. Users rely on the CS for cloud data storage and maintenance .They may also dynamically interact with the CS to access and update their stored data for various application purposes. The users may resort to TPA for ensuring the storage security of their outsourced data, while hoping to keep their data private from TPA. The Author considers the existence of a semi-trusted CS as does. Namely, in most of time it behaves properly and does not deviate from the prescribed protocol execution. However, during providing the cloud data storage based services, for their own benefits the CS might neglect to keep or deliberately delete rarely accessed data files which belong to ordinary cloud users. Moreover, the CS may decide to hide the data corruptions caused by server hacks or Byzantine failures to maintain reputation. It is assumed that the TPA, who is in the business of, auditing, is reliable and independent, and thus has no incentive to collude with either the CS or the users during the auditing process. TPA should be able to efficiently audit the cloud data storage without local copy of data and without bringing in additional on-line burden to cloud users. The Cloud Computing model of computing is a distributed application structure that partitions tasks or workloads between the providers of a resource or service, called Cloud servers, and service requesters, called clients [3].

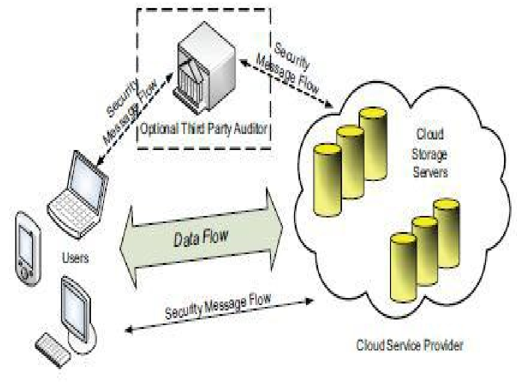

Fig 3: system model

3.4 Achieving Secure, Scalable, And Fine-Grained Data Access Control In Cloud Computing

The proposed scheme enables the data owner to delegate tasks of data file re-encryption and user secret key update to cloud servers without disclosing data contents or user access privilege information .Author achieves this goal by exploiting and uniquely combining techniques of attribute-based encryption (ABE), proxy re-encryption, and lazy re-encryption. The proposed scheme also has salient properties of user access privilege confidentiality and user secret key accountability and achieves fine graininess, scalability and data confidentiality for data access control in cloud computing. Extensive analysis shows that this proposed scheme is highly efficient and provably secures under existing security models [9].
Advantages
Low initial capital investment
Shorter start-up time for new services
Lower maintenance and operation costs
Higher utilization through virtualization
Easier disaster recovery
In order to achieve secure, scalable and fine-grained access control on outsourced data in the cloud, the author utilize and uniquely combine the following three advanced cryptographic techniques:
Key Policy Attribute-Based Encryption (KP-ABE).
Proxy Re-Encryption (PRE)
Lazy re-encryption
Module Description
Key Policy Attribute-Based Encryption (KP-ABE):
KP-ABE is a public key cryptography primitive for one-to-many communications. In KP-ABE, data are associated with attributes for each of which a public key component is defined. User secret key is defined to reflect the access structure so that the user is able to decrypt a cipher text if and only if the data attributes satisfy his access structure. A KP-ABE scheme is composed of four algorithms which can are listed as follows:
Setup Attributes
Encryption
Secret key generation
Decryption
Proxy Re-Encryption (PRE):
Proxy Re-Encryption (PRE) is a cryptographic primitive in which a semi-trusted proxy is able to convert a cipher text encrypted under Alice's public key into another cipher text that can be opened by Bob's private key without seeing the underlying plaintext.
Lazy re-encryption:
The lazy re-encryption technique allows Cloud Servers to aggregate computation tasks of multiple operations. The operations such as:
Update secret keys
Update user attributes.

3.5 Toward Publicly Auditable Secure Cloud Data Storage Services

The authors propose that publicly auditable cloud data storage is able to help this nascent cloud economy become fully established. With public auditability, a trusted entity with expertise and capabilities data owners do not possess can be delegated as an external audit party to assess the risk of outsourced data when needed. Such an auditing service not only helps save data owners' computation resources but also provides a transparent yet cost-effective method for data owners to gain trust in the cloud. The author describe approaches and system requirements that should be brought into consideration, and outline challenges that need to be resolved for such a publicly auditable secure cloud storage service to become a reality [10].

3.6 Online Data Storage Using Implicit Security

The authors describe the use of a data partitioning scheme for implementing such security involving the roots of a polynomial in finite field. The partitions are stored on randomly chosen servers on the network and they need to be retrieved to recreate the original data. Data reconstruction requires access to each server, login password and the knowledge of the servers on which the partitions are stored. This scheme may also be used for data security in sensor networks and internet voting protocols. The authors have described an implicit security architecture suited for the application of online storage. In this scheme data is partitioned in such a way that each partition is implicitly secure and does not need to be encrypted. These partitions are stored on different servers on the network which are known only to the user. Reconstruction of the data requires access to each server and the knowledge as to which servers the data partitions are stored. Several variations of this scheme are described, which include the implicit storage of encryption keys rather than the data, and where a subset of the partitions may be brought together to recreate the data [11].

3.7 Identity-Based Authentication for Cloud Computing

The authors propose an identity-based encryption (IBE) and decryption and identity-based signature (IBS) schemes for IBHMCC. Based on the former IBE and IBS schemes, an identity based authentication for cloud computing (IBACC) is proposed. The author presented an identity based authentication for cloud computing, based on the identity-based hierarchical model for cloud computing (IBHMCC) and corresponding encryption and signature schemes. They have proposed Identity-based Authentication Protocol. Identity-based Authentication Protocol contains sequence of steps. In step (1), the client C sends the server S a ClientHello message. The message contains a fresh random number C n, session identifier ID and C specification. In step (2), the server S responds with a ServerHello message which contains a new fresh random number S n, the session identifier ID and the cipher specification S specification The ciphertext is transmitted to C as ServerKeyExchange message. Then S generates a signature Sig S S [M] as the IdentityVerify message to forward to C. Finally, The ServerHelloDone message means the step (2) is over. In step (3), C firstly verifies the signature S Sig S S with the help of S ID Being certificate-free, the authentication protocol aligned well with demands of cloud computing. Performance analysis indicated that the authentication protocol is more efficient and lightweight than SAP, especially the more lightweight user side [12].

3.8 Security Framework Of Cloud Data Storage Based On Multi Agent System Architecture

The authors proposes Multi-Agent System (MAS) techniques that can be beneficial in cloud computing platform to facilitate security of cloud data storage (CDS) among it.MAS architecture offered eleven security attributes generated from four main security policies of correctness, integrity, confidentially and availability of users' data in the cloud.

3.9 Privacy-Preserving Public Auditing For Secure Cloud Storage

A Public Auditing Scheme Consists of four algorithms (KeyGen, SigGen, GenProof, VerifyProof)
KeyGen: key generation algorithm that is run by the user to setup the scheme
SigGen: used by the user to generate verification metadata, which may consist of MAC, signatures or other information used for auditing.
GenProof: run by the cloud server to generate a proof of data storage correctness
VerifyProof: run by the TPA to audit the proof from the cloud server. The author uses homomorphic authenticator technique for aggregate the data. Also uses a random mask technique achieved by a Pseudo Random Function (PRF)

4 CONCLUSION

The way cloud has been dominating the IT market, a major-shift towards the cloud can be expected in the coming years thus data security on the cloud would be the major concern for all the service providers. My survey consists of various existing data storage security techniques for cloud computing.

REFERENCES

[1] K.Valli Madhavi, R.Tamilkodi, R.BalaDinakar, –Data Storage Security in Cloud Computing for Ensuring Effective and Flexible Distributed System‖, International Journal of Electronics Communication and Computer Engineering, 2012.

[2] K. Kajendran, J.James Jeyaseelan J. Jakkulin Joshi, –An Approach For Secured Data Storage Using Cloud Computing‖, International Journal of Computer Trends and Technology, 2011.

[3] Cong Wang, Qian Wang, Kui Ren , Wenjing Lou, –Ensuring Data Storage Security in Cloud Computing‖.

[4] B. Shwetha Bindu1, B. Yadaiah, –Secure Data Storage In Cloud Computing‖, International Journal of Research in Computer Science, 2011.

[5] Nilesh N. Kumbhar, Virendrasingh V. Chaudhari, Mohit A.Badhe,

–The Comprehensive Approach for Data Security in Cloud Computing: A Survey‖, International Journal of Computer Applications (0975 – 8887), 2012.

[6] Rohit Bhadauria, Rituparna Chaki, Nabendu Chaki, Sugata Sanyal, –A Survey on Security Issues in Cloud Computing‖.

[7] S.Sajithabanu, Dr.E.George Prakash Raj, –Data Storage Security in Cloud‖, 2011.

[8] Zoran Pantic and Muhammad Ali Babar, –Guidelines for Building a Private Cloud Infrastructure,‖ in IT University of Copenhagen, Denmark, 2012.

[9] Shucheng Yu., Cong Wan, Kui Ren, Wenjing Lou.,"Achieving Secure, Scalable, and Fine-grained Data Access Control in Cloud Computing‖, IEEE Communications Society for publication,2010.

[10] Qian Wang, Cong Wang, Kui Ren, Wenjing Lou, Jin Li,"Toward Publicly Auditable Secure Cloud Data Storage Services‖, IEEE Network, 2010.

[11] Abhishek Parakh, Subhash Kak, "Online data storage using implicit security‖, 2009.

[12] Hongwei Li, Yuanshun Dai, Ling Tian, Haomiao Yang,"Identity-Based Authentication for Cloud Computing‖, CloudCom 2009, 2009.

[13] Sushil Bhardwaj, Leena Jain and Sandeep Jain,‖ cloud comp uting: a study of infrastructure as a service (IAAS),‖ M.M.University, Mullana (Ambala, India) 133203, 2010.

[14] Mladen A. Vouk, –Cloud Computing – Issues,Research and Implementations‖,2008.