A Study On Data Security Issues In Public Cloud

Alycia Sebastian, Dr. L. Arockiam

Abstract: The cloud computing concept has been evolving for more than 40 years. Cloud computing is an on demand computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. The cloud computing services are delivered through software as service (SaaS), platform as service (PaaS) and Infrastructure as service (IaaS). Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services are provided by cloud service providers. The common data security concerns are securing data in transit and at rest, access control and data separation. In this survey paper, we review the data in public cloud, identify and discuss the security risks associated with it and analyze its solution strategies.

Keyword: data security, encryption, Key management, public cloud.

I. Introduction

Moving to cloud is the growing trend in many enterprises to cut down the cost in establishing an IT infrastructure. The Cloud Computing is an Internet based computing, which offers resources such as servers, data storage systems and applications as “pay per use” service in real time over internet. Cloud computing allows user to access data and application from anywhere at any time, reduces hardware cost and allows SME’s to save money on IT support by paying metered fee to service providers. A Working Definition of Cloud Computing of National Institute of Standards and Technology (NIST)[1] is as follows: Cloud computing is 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. Cloud services are categorized into SaaS (Software as Service), Paas (Platform as a Service) and IaaS (Infrastructure as a Service). The bottom layer IaaS provides services such as storage, virtualization management, Networks and Servers. The intermediate PaaS layer basic functionality is to provide a platform for developing applications and it offers services such as database management system, application development tools and web server. The SaaS layer allows users to use tha pplications such as email, ERP, CRM on pay-per use basis.

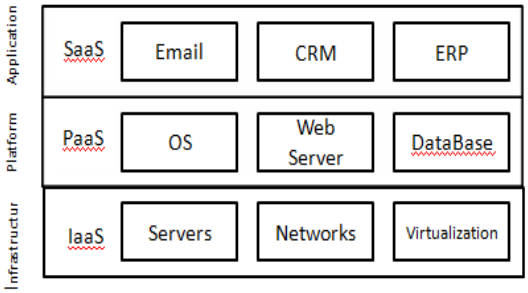

Figure: Cloud Service Layers

- Alycia Sebastian, Waljat college of Applied Sciences Muscat, Oman alycia.sebastian@gmail.com
- Dr. L. Arockiam, St. Joseph’s College (Autonomous) Tiruchirappalli, TamilNadu, India larockiam@yahoo.co.in

The cloud itself is classified into public or private cloud [2]. Public cloud is access to anyone on the internet. The services may be free or pay for what you use. Private cloud is limited to people in an organization and the cloud is managed and controlled by the organization itself. From small scaled companies to larger companies and virtually in every field, people are now shifting to cloud computing to avoid the overhead caused by traditional applications. Cloud computing reduces the complexity of handling large volumes of data, provide higher processing speed, increased scalability, storage and access to new technologies. Cloud extends the existing resources to meet the organization needs, so it has become a necessity to move to cloud computing. Cloud computing is divided into two major sections: Front end and back end. Front end is the client machine and back end is the cloud which comprises of computers, servers and storage machines [3]. The public cloud is owned and managed by cloud service providers (CSP). The major CSP’s are Amazon, Rackspace, Verizon and Microsoft. The user data is stored in the datacenter of the providers and are managed by the providers. The user’s personal data are normally processed in the cloud. The security of the data in storage and during transcation has become a major concern in implementing cloud in an organization. Also the CSP can rapidly transfer the personal data and information from one datacenter to another, so another major concern is that the owner usually has no control or knowledge about the exact location of the provided resources.

II. Understanding Data security issues and analyzing its existing solutions in public Cloud

In a recent survey [4] conducted, 69% among 127 cloud providers and 35% cloud users see cloud users responsible for ensuring the security of the cloud services, 32% of the cloud providers and 32% of the cloud users see security as the cloud providers responsibility, 16% of the cloud providers and 33% of the cloud users see it as a shared responsibility.

A. Encryption

The decision to move to cloud computing from traditional IT infrastructure results in higher number of security breaches. The increase in the data transfer between different business units of an organization has heightened the risk of data security. Cloud is a distributed network. Data may be distributed across the world. Organizations at any point will have no knowledge or control over the location of data

storage. So to protect data, encrypting the data and controlling the encryption keys is the only solution. To give maximum security, data should be encrypted at the source point, then during transmission and in the cloud. Depending on the providers to offer security is a risk that no organization could afford. So, encrypting the data locally before transmitting to cloud solely rests on the customers. Various symmetric and asymmetric encryption algorithms are in use and its security level depends on the key size. Symmetric algorithms use a preshared key between sender and receiver. Common symmetric algorithms used are AES which uses key size 128 bits (minimum) to 256-bits for highly sensitive data, DES with 56-bit keys and Triple DES which gives 112 bits of security with 168-bit keys [5]. In [6] the authors have compared the symmetric algorithms and have found AES better option since it produces less overhead and is faster [7] than DES and Triple DES. Public key algorithms use two separate keys, public and private keys. Common asymmetric algorithms are RSA, Diffie-Helman and ECC. 2048-bits RSA and DH both provide the same security strength as that of 112-bits symmetric keys[6]. Asymmetric algorithms are not suitable for encrypting large content since it is computationally costlier than symmetric algorithms because of its larger keys [8]. These algorithms are used in secure distribution of preshared keys and symmetric algorithms are then used in encryption.

B. Homomorphic Encrytpion

For healthier security, it is better to encrypt data before passing it to the cloud. But client side encryption leads to technical challenge since data processing or query on encrypted database is not possible. The vendor needs the client's private key to decrypt the data for processing each time. This affects the important aspect of cloud computing i.e. confidentiality of the data stored in the cloud. IBM researcher Craig Gentry's breakthrough work in 2009, proposes the first fully Homomorphic Encryption (FHE) [9] scheme capable of performing arbitrary number of additions and multiplications on encrypted data without being decrypted. The vendors can perform computation on the encrypted data using FHE for client and the result can be decrypted only by the client. Since then many research papers have been published on homomorphic encryption. In paper [10] the authors have presented a fully homomorphic encryption scheme which produces short cipertext and its security is based on the learning with errors (LWE) assumptions. The proposed solution is based on hardness of problems on arbitrary lattices while Gentry's work was based on ideal lattices. Homomorphic encryption requires intense computational requirements and also it affects cloud efficiency, so implementing this scheme fully in cloud is still in infancy stage.

C. Key Management

Cloud data should be protected from data loss and theft. Encrypting the data in the cloud is the only possible solution. The strength of the encryption algorithms depends on the security of the encryption keys. Key management is the important part in encryption. Key management involves generating, using, storing, distributing, revoking, verifying and destroying of keys [11]. Three options for Keys storage and management are [12]: in the company's datacenter which is the most securable one, SaaS Key management where the SaaS vendors are responsible for the security of the key and IaaS key management where the key is given to IaaS vendors for safeguarding. Of all the three approaches, key management by the customer is the best solution. Various Key management approaches which are in practice are using pre-shared keys, public key cryptography algorithms and Key distribution centers [13]. Cloud providers must deploy a secure mechanism for key management in the cloud. But it's best if the key management is handled by data owner for stronger security.

D. Data Transaction

Secure Socket Layer (SSL) provides secure data communication over unsecured network. SSL protocol uses public key encryption for key exchange, symmetric encryption for confidentiality and message authentication code for message integrity [14]. Several SSL certificate companies including DigiNotar and Comoda [15] were compromised last year to create fraudulent certificates for sites like Google, Yahoo, and Microsoft's Hotmail. So another option is to use DNSSEC to authenticate the origin and integrity of DNS data as it traverses the Internet. DNSSEC is not an encryption protocol, rather an additional security at the DNS level. DNSSEC is a form of public key infrastructure (PKI) [16]. The drawback is that DNSSEC does not validate the authenticity of the website that the user seeks. The solution is to combine SSL and DNSSESC [17] to provide a secure and authenticated data communication

E. Authentication

Since public cloud is an open platform, anyone with internet can access the data. The important question is: which user can use and view the data in the cloud? A stronger authentication approach is needed to ensure that user data is protected from unwanted access. Cloud experts say, organizations or consumers should start using Multi Factor authentication to protect their data from hacking [18]. US Federal regulators recognize three authentication factors: password or PIN, smart card, PKI certificate, USB key or mobile number and biometric characteristics like fingerprint or voice pattern [19]. Multi Factor authentication system is to authenticate users with two or more combination with the above factors.

F. Data Persistence

As more and more companies are moving towards cloud computing, the research reveals still there is lack of maturity in cloud service market. The major issue in data storage is data persistence. The cloud users may face the risk of exposure of residual data when they do data removal or data transfer. CSP offers no standardized features of how securely the disk space or memory is recycled. A recent research by Context Information Security [20] revealed a serious data security flaws in Amazon EC2, Gigenet, Rackspace and VPS.net. All four CSP's VM lacked up-to-date security patches and did not have antivirus software included in it. The most serious flaw from Rackspace and VPS.Net, was data remanance. Rackspace has fixed the problem by moving all its Linux based VM customers to Citrix XenServer and VPS.Net has also confirmed its fix. Customer must know what they are getting before adopting to cloud computing. Even after moving data from own datacenter to cloud database, it is the customer

responsibility to identify the critical data and how it is used and protected.

G. Data storage

User's data in the cloud is maintained in centralized database which are owned by the providers. Cloud databases offer significant advantages including automatic failover and recovery, load balancing, automatic backups and optimal auto scaling [21]. Users can either run database independently or use database as a service provided by CSP. Database in the cloud can be SQL database such as Oracle Database, Microsoft SQL Server and MySQL or NOSQL database such as Apache Cassandra, CouchDB and MongoDB [22]. Another important challenge is encrypting the archived data.

III Conclusion

Cloud computing eventhough a promising technology, concern arises in terms of security and privacy. The customers before shifting their datacenter to cloud, they should understand the sensitivity of data and choose the cloud model based on their requirement. Though the CSP plays the major role in providing security for the data deployed in the cloud, the customers should know what security system the providers use and also obtain the ownership of the data i.e. to use and manage their own data. It is the shared responsibility of cloud providers and customers to maintain a secure environment.

References

[1]. Peter Mell and Tim Grance. "The NIST Definition of Cloud Computing." National Institute of Standards and Technology, Information Technology Laboratory, September 2011. http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

[2]. http://searchcloudcomputing.techtarget.com/definition/cloud-computing

[3]. http://computer.howstuffworks.com/cloud-computing-crash-course.htm

[4]. 9 May 2011, http://elastic-security.com/2011/05/09/cloud-security-who-is-responsible/

[5]. http://en.wikipedia.org/wiki/Key_size,http://www.nsa.gov/business/programs/elliptic_curve.shtml

[6]. Krunal Suthar, Parmalik Kumar, Hitesh Gupta, Hiren Patel, "Analytical Comparison of Symmetric Encryption and Encoding Techniques for Cloud Environment", International Journal of Computer Applications (0975 – 8887), Volume 60– No.19, December 2012

[7]. Hirani, "Energy Consumption of Encryption schemes in wireless device" Thesis, university of Pittsburgh, Retrieved Oct.1, 2008.

[8]. http://en.wikipedia.org/wiki/Public-key_cryptography#Weaknesses

[9]. Craig Gentry ,"A Fully Homomorphic Encryption scheme", A dissertation submitted to the department of computer science and the committee on graduate studies of stanford university in partial fulfillment of the requirements for the degree of doctor of philosophy, September 2009.

[10]. Zvika Brakerski, Vinod Vaikuntanathany,"Efficient Fully Homomorphic Encryption from (Standard) LWE" , FOCS 2011.

[11]. Lee Badger, Cloud Computing: Some Implications for Key Management, http://csrc.nist.gov/groups/ST/key_mgmt/index.html , June 2009

[12]. Todd Thieman ,"https://blog.cloudsecurityalliance.org/2012/03/12/lock-box-where-should-you-store-cloud-encryption-keys/, March 12, 2012

[13]. Elaine Barker,Dennis Branstad ,Santosh Chokhani ,Miles Smid, Cryptographic Key Management Workshop Summary – June 8-9, 2009

[14]. http://en.wikipedia.org/wiki/Transport_Layer_Security

[15]. Michael A. Davis, 2012 Data Encryption Survey: Progress And Pain http://www.informationweek.com/security/encryption/2012-data-encryption-survey-progress-and/232500062, January 30, 2012

[16]. http://www.icann.org/en/about/learning/factsheets/dnssec-qaa-09oct08-en.htm

[17]. http://www.symantec.com/connect/blogs/dnssec-and-ssl-better-together

[18]. Sean Ludwig, cloudbeat2012, 28 November 2012, http://venturebeat.com/2012/11/28/cloud-security-cloudbeat-2012

[19]. http://www.onelogin.com/product/strong-authentication/two-factor-authentication/

[20]. http://www.computerweekly.com/news/2240148943/Investigation-reveals-serious-cloud-computing-data-security-flaws

[21]. http://www.webopedia.com/TERM/C/cloud_database.html

[22]. http://en.wikipedia.org/wiki/Cloud_database