ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 3, March 2012

Survey on Data Security Issues and Data Security Models in Cloud Computing

Ramasami S., Umamaheswari P.

***Abstract*— with the rapid development of cloud computing, providing security to the customer's data becomes more and more important. The customer doesn't know where their data has been stored in the cloud. This paper analyses the various data storage security issues and data privacy issues. The issue includes are Insecure Interfaces and APIs, Service Hijacking, Malicious Insiders, Unknown Risk Profile, potential data breaches, data leakage and interception. Next we presented a security model for providing data security using traditional and homomorphic techniques and finally we gave a technique for preserving data privacy.**

Index Terms—**Cloud computing, Data security, Data privacy, Homomorphic algorithm.**

I. INTRODUCTION

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction through internet. The Internet cloud works as a service factory built around virtualized data centers. Cloud platforms are dynamically built through virtualization with provisioned hardware, software, networks, and datasets. The idea is to migrate desktop computing to a service-oriented platform using virtual server clusters at data centers. However, a lack of trust between cloud users and providers has hindered the universal acceptance of clouds as outsourced computing services [1] [2].

II. CLOUD COMPUTING SERVICE MODELS

In cloud computing, everything is delivered as a Service (XaaS), from testing and security, to collaboration and metamodeling [8]. The cloud was rapidly becoming a conflagration of buzzwords "as a service". Today there are three main service models, which are agreed on and defined in the NIST document.

1. ***Software as a Service {SaaS}*** - this simply means delivering software over the Internet. It is the most widely known model of cloud computing. *SaaS* has been around since early 2001 when it was commonly referred to as the Application Service Provider (ASP) Model [8]. Software as a Service consists of software running on the provider's cloud infrastructure, delivered to (multiple) clients (on demand) via a thin client (e.g. browser) over the Internet. Typical examples are Google Docs and Salesforce.com CRM.
2. ***Platform as a Service {PaaS}*** - this gives a client (developer) the flexibility to build (develop, test and deploy) applications on the provider's platform (API, storage and infrastructure). PaaS stakeholders include the PaaS hoster who provides the infrastructure (servers etc), the PaaS provider who provides the development tools and platform and the PaaS user [10]. Examples of PaaS are Microsoft Azure and Google AppEngine.
3. ***Infrastructure as a Service {IaaS}*** - rather than buy servers and build a datacenter from ground up, and consequently having to worry about what happens when the website hits a million users, IaaS offers users elastic on demand access to resources (networking, servers and storage), which could be accessed via a service API. The underlying infrastructure is transparent to the end user, while s/he retains control over the platform and software running on the infrastructure. IaaS runs on a tenancy model, which employs a usage-based payment approach allowing users to pay for only those resources they actually use [4].

III. CLOUD COMPUTING DEPLOYMENT MODELS

Depending on infrastructure ownership, there are four deployment models of cloud computing each with its merits and demerits. This is where the security issues start.

1. ***The Public Cloud* -** this is the traditional view of cloud computing in every day lingua. It is usually owned by a large organization (e.g. Amazon's EC2, Google's AppEngine and Microsoft's Azure). The owner-organization makes its infrastructure available to the general public via a multi-tenant model on a self-service basis delivered over the Internet. This is the most cost-effective model leading to substantial savings for the user, albeit with attendant privacy and security issues since the physical location of the provider's infrastructure usually traverses numerous national boundaries [3].
2. ***The Private Cloud* -** refers to cloud infrastructure in a single tenant environment. It defers from the traditional datacenter in its predominant use of virtualization. It may be managed by the tenant organization or by a third party within or outside the tenant premises. A private cloud costs more than the public cloud, but it leads to more cost savings when compared with a datacenter as evidenced by Concur Technologies (est. savings of $7 million in 3 years from 2009) [11]. The private cloud gives an organization greater control over its data and resources. As a result, the private cloud is more appealing to enterprises especially in mission and safety critical organizations.
3. ***The Community Cloud* -** according to NIST, the community cloud refers to a cloud infrastructure shared

by several organizations within a specific community. It may be managed by any one of the organizations or a third party. A typical example is the Open Cirrus Cloud Computing Tested, which is a collection of Federated data centers across six sites spanning from North America to Asia [12].

IV. Data Security Issues

A. *Abuse and Nefarious Use of Cloud Computing*

IaaS providers offer their customers the illusion of unlimited compute, network, and storage capacity often coupled with a 'frictionless' registration process where anyone with a valid credit card can register and immediately begin using cloud services. Some providers even offer free limited trial periods. By abusing the relative anonymity behind these registration and usage models, spammers, malicious code authors, and other criminals have been able to conduct their activities with relative impunity. PaaS providers have traditionally suffered most from this kind of attacks; however, recent evidence shows that hackers have begun to target IaaS vendors as well. Future areas of concern include password and key cracking, DDOS, launching dynamic attack points, hosting malicious data, botnet command and control, building rainbow tables, and CAPTCHA solving farms.

B. *Shared Technology Issues*

IaaS vendors deliver their services in a scalable way by sharing infrastructure. Often, the underlying components that make up this infrastructure (e.g., CPU caches, GPUs, etc.) were not designed to offer strong isolation properties for a multi-tenant architecture. To address this gap, a virtualization hyper visor mediates access between guest operating systems and the physical compute resources. Still, even hyper visors have exhibited flaws that have enabled guest operating systems to gain inappropriate levels of control or influence on the underlying platform. A defense in depth strategy is recommended, and should include compute, storage, and network security enforcement and monitoring. Strong compartmentalization should be employed to ensure that individual customers do not impact the operations of other tenants running on the same cloud provider. Customers should not have access to any other tenant's actual or residual data, network traffic, etc [5] [6].

C. *Account or Service Hijacking:*

Account or service hijacking is not new. Attack methods such as phishing, fraud, and exploitation of software vulnerabilities still achieve results. Credentials and passwords are often reused, which amplifies the impact of such attacks. Cloud solutions add a new threat to the landscape. If an attacker gains access to your credentials, they can eavesdrop on your activities and transactions, manipulate data, return falsified information, and redirect your clients to illegitimate sites. Your account or service instances may become a new base for the attacker. From here, they may leverage the power of your reputation to launch subsequent attacks.

D. *Unknown Risk Profile:*

One of the tenets of Cloud Computing is the reduction of hardware and software ownership and maintenance to allow companies to focus on their core business strengths. This has clear financial and operational benefits, which must be weighed carefully against the contradictory security concerns complicated by the fact that cloud deployments are driven by anticipated benefits, by groups who may lose track of the security ramifications. Versions of software, code updates, security practices, vulnerability profiles, intrusion attempts, and security design, are all important factors for estimating your company's security posture. Information about who is sharing your infrastructure may be pertinent, in addition to network intrusion logs, redirection attempts and/or successes, and other logs. Security by obscurity may be low effort, but it can result in unknown exposures. It may also impair the in-depth analysis required highly controlled or regulated operational areas [7].

V. Data Privacy Issues

A. *Insecure Interfaces and APIs:*

Cloud computing providers expose a set of software interfaces or APIs that customers use to manage and interact with cloud services. Provisioning, management, orchestration, and monitoring are all performed using these interfaces.

The security and availability of general cloud services is dependent upon the security of these basic APIs. From authentication and access control to encryption and activity monitoring, these interfaces must be designed to protect against both accidental and malicious attempts to circumvent policy. Furthermore, organizations and third parties often build upon these interfaces to offer value-added services to their customers. This introduces the complexity of the new layered API; it also increases risk, as organizations may be required to relinquish their credentials to third parties in order to enable their agency.

B. *Malicious Insiders:*

The threat of a malicious insider is well-known to most organizations. This threat is amplified for consumers of cloud services by the convergence of IT services and customers under a single management domain, combined with a general lack of transparency into provider process and procedure. For example, a provider may not reveal how it grants employees access to physical and virtual assets, how it monitors these employees, or how it analyzes and reports on policy compliance [8].

To complicate matters, there is often little or no visibility into the hiring standards and practices for cloud employees. This kind of situation clearly creates an attractive opportunity for an adversary ranging from the hobbyist hacker, to organized crime, to corporate espionage, or even nation-state sponsored intrusion. The level of access granted could enable such an adversary to harvest confidential data or gain complete control over the cloud services with little or no risk of detection.

C. *Data Loss or Leakage:*

There are many ways to compromise data. Deletion or alteration of records without a backup of the original content is an obvious example. Unlinking a record from a larger context may render it unrecoverable, as can storage on unreliable media. Loss of an encoding key may result in effective destruction. Finally, unauthorized parties must be prevented from gaining access to sensitive data. The threat of data compromise increases in the cloud, due to the number of and interactions between risks and challenges which are either unique to cloud, or more dangerous because of the architectural or operational characteristics of the cloud environment [9].

VI. Security Model

The network representative architecture for cloud data storage, which contains three parts as shown in Figure 2, viz Users, Cloud Service Provider (CSP) and Third Party Auditor (TPA).

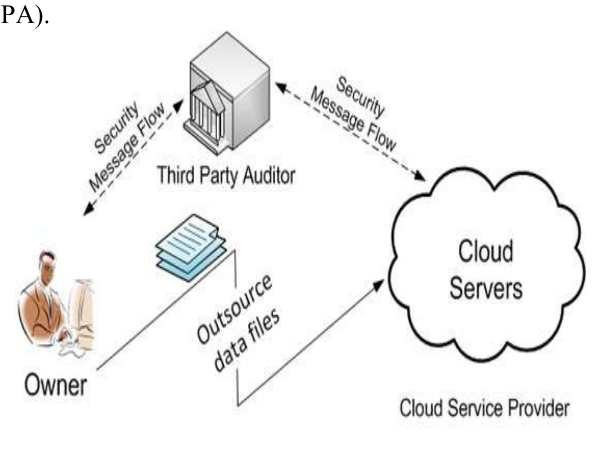

Fig. 1. Cloud Data Storage Architecture

As shown in Figure 2, the brief descriptions of these parts as follows:

(a) Users: - Users who have data to be stored and interact with the cloud service provider (CSP) to manage their data on the cloud. They are typically, desktop computers, laptops, tablet computers, mobile phones, etc.

(b) Cloud Service Provider (CSP):- Cloud service provider (CSP) has major resources and expertise in building and managing distributed cloud storage servers. A CSP offers storage or software services to user's available via the Internet [10].

(c) Third Parity Auditor (TPA):- An optional TPA, who has expertise and capabilities that users may not have, is monitoring the risk of cloud data storage services on behalf of users.

A. *Traditional Methods:*

A straightforward approach to protect the data integrity would be using traditional cryptographic methods, such as the well-known message authentication codes (MACs). Initially, data owners can locally maintain a small amount of MACs for the data files to be outsourced. Whenever the data owner needs to retrieve the file, she can verify the integrity by recalculating the MAC of the received data file and comparing it to the locally pre computed value. If the data file is large, a hash tree [11] can be employed, where the leaves are hashes of data blocks and internal nodes are hashes of their children of the tree. The data owner only needs to store the root hash of the tree to authenticate his received data.

While this method allows data owners to verify the correctness of the received data from the cloud, it does not give any assurance about the correctness of other outsourced data. In other words, it does not give any guarantee that the data in the cloud are all actually intact, unless the data are all downloaded by the owner. Because the amount of cloud data can be huge, it would be quite impractical for a data owner to retrieve all of her data just in order to verify the data is still correct. If the data auditing task is delegated to a TPA, this method inevitably violates our suggested requirements, with large auditing cost for a cloud server (for accessing and transferring all of the data) and data privacy exposure to the TPA (for retrieving a local copy of data). Thus, new approaches are required.

To avoid retrieving data from the cloud server, a simple improvement to this straightforward solution can be performed as follows: Before data outsourcing, the owner chooses a set of random MAC keys, pre computes the MACs for the whole data file, and publishes these verification metadata to the TPA. The TPA can each time reveal a secret MAC key to the cloud server and ask for a fresh keyed MAC for comparison. In this way the bandwidth cost for each audit is only at the bit-length level (keys and MACs). However, a particular drawback is that the number of times a data file can be audited is limited by the number of secret keys that must be fixed a priori, which might introduce an additional online burden to the data owner: Once all possible secret keys are exhausted, the data owner then has to retrieve data from the server in order to recompute and republish new MACs to the TPA. Another drawback of this improved approach is its inability to deal with data dynamics, as any data change would make those precomputed MACs completely unusable [13].

B. *Utilizing Homomorphic Authenticators*

To significantly reduce the arbitrarily large communication overhead for public auditability without introducing any online burden on the data owner, we resort to the homomorphic authenticator technique [7, 10]. Homomorphic authenticators are unforgivable metadata generated from individual data blocks, which can be securely aggregated in such a way to assure a verifier that a linear combination of data blocks is correctly computed by verifying only the aggregated authenticator. Using this technique requires additional information encoded along with the data before outsourcing. Specifically, a data file is divided into n blocks mi (i =1,…, n), and each block mi has a corresponding homomorphic authenticator σi computed as its metadata to ensure the integrity.

Every time it must be verified that the cloud server is honestly storing the data, the data owner or TPA can submit challenges chal = $\{(i, \nu_i)\}$ for sampling a set of randomly selected blocks, where $\{\nu_i\}$ can be arbitrary weights. Due to the nice property of the homomorphic authenticator, server only needs to response a linear combination of the sampled data blocks $\mu = \Sigma_i v_i.m_i$, as well as an aggregated authenticator $\sigma = \Pi_i \sigma_i^{vi}$, both computed from $\{m_i, \sigma_i, v_i\}_{i€chal}$. Once the response of μ and σ is verified by TPA, then high probabilistic guarantee on large fraction of cloud data correctness can be obtained.1 Because off-the-shelf error-correcting code technique can be adopted before data outsourcing [6, 10], large fraction of correct cloud data would be sufficient to recover the whole data. Note that for typical choices of block size │mi│ and file block number n, where │*mi*│ *>> log (n),* the response μ and σ are (essentially) about the same size as individual block mi and σ_i. This means almost constant communication overhead, independent of file size, for each auditing can be achieved. Moreover, since the TPA could regenerate the fresh random sampling challenges, unbounded auditing is achieved too, which means no additional on-line burden would be incurred towards data owner. However, despite the desirable properties, this approach only works well for encrypted data. When directly applied to unencrypted data, it still leaks bits information towards TPA, as discussed next.

VII. Protecting Data Privacy

Homomorphic authenticator technique is used to provide the data privacy in the cloud. From the perspective of protecting data privacy, the owners, who own the data and rely on the TPA just for the storage security of their data, do not want this auditing process introducing new vulnerabilities of unauthorized information leakage into their data security [12],[14]. Moreover, there are legal regulations, such as the U.S. Health Insurance Portability and Accountability Act (HIPAA), further demanding the outsourced data not to be leaked to external parties. Exploiting data encryption before outsourcing [11],[12] is one way to mitigate this privacy concern, but it is only complementary to the privacy-preserving public auditing scheme to be deployed in cloud.

Without a properly designed auditing protocol, encryption itself cannot prevent data from flowing away toward external parties during the auditing process. Thus, it does not completely solve the problem of protecting data privacy but just reduces it to the one of managing the encryption keys. Unauthorized data leakage still remains a problem due to the potential exposure of encryption keys. To address this concern, a proper approach is to combine the homomorphic authenticator with random masking. This way, the linear combination of sampled blocks in the server's response is masked with randomness generated by the server. With random masking, the TPA no longer has all the necessary information to build up a correct group of linear equations and therefore cannot derive the owner's data content, no matter how many linear combinations of the same set of file blocks can be collected. Meanwhile, due to the algebraic property of the homomorphic authenticator, the correctness validation of the block-authenticator pairs (μ and σ) can still be carried out in a new way, even in the presence of randomness. Some initial results of this branch of research can be found in [14].

This improved technique ensures the privacy of owner data content during the auditing process, regardless of whether or not the data is encrypted, which definitely provides more flexibility for different application scenarios of cloud data storage. Besides, with the homomorphic authenticator, the desirable property of constant communication overhead for the server's response during the audit is still preserved.

VIII. Conclusion

Cloud computing moves the application software and databases to servers in large data centers on the Internet, where the management of the data and services are not fully trustworthy. This unique attribute raises many new security challenges in areas such as software and data security, recovery, and privacy, as well as legal issues in areas such as regulatory compliance and auditing, all of which have not been well understood. In this article we focus on cloud data storage security. We first present a cloud service models, cloud service deployment model and various data security issues in the cloud. We then suggest a Security model for providing data security and providing data privacy in the cloud.

References

[1] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic."Cloud Computing and Emerging IT Platforms: Vision, Hype, and Reality for Delivering Computing as the 5th Utility," Future Generation Computer Systems, vol. 25, no. 6, June 2009, pp 599–616.

[2] H. Takabi, J.B.D. Joshi, and G. Ahn, "Security and Privacy Challenges in Cloud Computing Environments", Article in IEEE Security and Privacy, vol. 8, no.6, Nov- Dec. 2010, pp. 24-31.

[3] N. Gohring, "Amazon's S3 down for several hours," Online at http://www.pcworld.com /businesscenter/ articl/ 142549/amasons_down_for _sever_hours.html", 2008.

[4] Bhaskar P., Admela J·, Dimitrios K·, Yves G.:Architectural Requirements for Cloud Computing Systems: An Enterprise Cloud Approach. J. Grid Computing 9(1), 3- 26 (2011).

[5] Ateniese G, Kamara S, Katz J. Proofs of Storage from homomorphic identification protocols. In: Proc. of ASIACRYPT '09, 2009, pp. 319-333.

[6] Ateniese G, Pietro R D, Mancini L V, Tsudik G. Scalable and efficient provable data possession. In: Proc. of SecureComm '08, 2008, pp.1-10.

[7] Xiao D, Shu J, Chen K, Zheng W. A Practical Data Possession Checking Scheme for Networked Archival Storage. Journal of Computer Research and Development , 2009, 46(10) : 1660-1668.

[8] G.Ateniese et al., "Provable Data Possession at Untrusted Stores," Proc. ACM CCS '07, Oct. 2007, pp. 598–609.

[9] C. Erway et al., "Dynamic Provable Data Possession," Proc. ACM CCS '09, Nov. 2009, pp. 213–22.

[10] M. A. Shah et al., "Auditing to keep Online Storage Services Honest," Proc. USENIX HotOS '07, May 2007.

[11] R. Gellman, “Privacy in the clouds: Risks to privacy and confidentiality from cloud computing”, Prepared for the World Privacy Forum, online at http://www.world privacy forum. Org/pdf/WPF Cloud Privacy Report. PDF, Feb 2009.

[12] Q.Wang et al., “Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing,” Proc. ESORICS ‘09, Sept. 2009, pp. 355–70.

[13] C.Wang et al.,”Ensuring Data Storage Security in Cloud Computing,” Proc.IWQoS ‘09, July 2009, pp. 1–9.

[14] C. Wang et al., “Privacy-Preserving Public Auditing for Storage Security in Cloud Computing,” Proc. IEEE INFOCOM ‘10, Mar. 2010.