An Efficient Approach to prevent Data Breaches in Cloud

Nina Pearl Doe, Mensah Sitti, V. Suganya

Abstract— Cloud Computing is a computing paradigm shift where computing is moved away from personal computers or an individual server to a cloud of computers. Its flexibility, cost-effectiveness, and dynamically re-allocation of resources as per demand make it desirable. At an unprecedented pace, cloud computing has simultaneously transformed business and government, and created new security challenges such as data breaches, data loss, account hijacking and denial of service. Paramount among these security threats is data breaches. The proposed work is to prevent data breaching threat by way of providing user authentication through one-time password system and challenge response, risk assessment to identify and prevent possible risks, encryption using enhanced elliptic curve cryptography where a cryptographically secure random number generation is used to make the number unpredictable, data integrity using MD5 technique, and key management. Also a secure disposal of information and secure transmission of files between clouds are ensured. The platform for deployment of the application is Google App Engine.

Index Terms— Authentication, Cloud security issues, Elliptic curve cryptography, MD5, Risk assessment, Role Based Access Control, Secure Disposal.

1 INTRODUCTION

NIST defines cloud computing as: "cloud computing is a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable computing resources that can be rapidly positioned with minimal management effort or service provider interaction". Cloud computing is transforming information technology. As information and processes migrate to the cloud, it is transforming not only where computing is done, but, fundamentally, how it is done [1]. As increasingly more corporate and academic worlds invest in this technology, it will also drastically change it professionals' working environment.

1.1 Selection of Cloud Service Model

Software as a service (SaaS) is a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet. In the software on demand model, the provider gives customers network-based access to a single copy of an application created specifically for SaaS distribution.

Platform as a service (PaaS) is a category of cloud computing services that provides a computing platform and a solution stack as a service. In this model, the consumer creates the software using tools and/or libraries from the provider. The consumer also controls software deployment and configuration settings. The provider provides the networks, servers, storage, and other services.

- *Nina Pearl Doe is currently pursuing masters degree program in Computer Science and Engineering in Anna University. E-mail: nina4perl@gmail.com*
- *Mensah Sitti is currently pursuring masters degree program in Computer Science and Engineering at Anna University. E-mail: mensitti@gmail.com*
- *V. Suganya is currently pursuing PHD degree program in Computer Science and Engineering in Anna University, India. E-mail: sxviswa@gmail.com*

Infrastructure as a service (IaaS): In the most basic cloud-service model, providers of IaaS offer computers, physical or (more often) virtual machines, and other resources. IaaS clouds often offer additional resources such as a virtual-machine disk image library, raw (block) and file-based storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles.

1.2 Selection of cloud deployment model

Private cloud: Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally. Undertaking a private cloud project requires a significant level and degree of engagement to virtualize the business environment, and requires the organization to re-evaluate decisions about existing resources.

Public cloud: A cloud is called a 'Public cloud' when the services are rendered over a network that is open for public use. Generally, public cloud service providers like Amazon AWS, Microsoft and Google own and operate the infrastructure and offer access only via Internet (direct connectivity is not offered).

Community cloud: Community cloud shares infrastructure be-tween several organizations from a specific community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted in-ternally or externally. The costs are spread over fewer users than a public cloud (but more than a private cloud), so only some of the cost savings potential of cloud computing are realized.

Hybrid cloud: Hybrid cloud is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models. Such composition expands deployment options for cloud services, allowing IT organizations to use public cloud computing resources to meet temporary needs.

Cloud computing, as a new technology, has also created new security challenges such as data breaches, data loss, account hijacking and denial of service. The shift from server to service-based thinking is transforming the way technology departments think about, design, and deliver computing technology and applications. Yet these advances have created new security vulnerabilities, including security issues whose full impact is still emerging. Among the most significant security risks associated with cloud computing is data breaching.

1.3 Data security issues in cloud computing

Compared with traditional software architecture, cloud compu-ting has more serious data security problem. Data security is aimed at applying technical mechanism to guarantee data management in reasonable control, and guarantee data without illegal visit or revise during data process. As mentioned above, the main data security problems of cloud computing include data breaches, data loss, account hijacking and denial of service. Our work aims at data breaching which is currently the biggest threat in cloud.

Data breaching is the biggest security issue [2]. A data breach is the intentional or unintentional release of secure information to an untrusted environment. It is a security incident in which sensitive, protected or confidential data is copied, transmitted, viewed, stolen or used by an individual unauthorized to do so. A capable hacker can easily get into a client side application and get into the client's confidential data. When a cloud customer puts its sensitive data into the cloud it is completely reliant on the security and incident response processes of the cloud service provider in order to respond to a data breach. This situation poses many fundamental problems.

When an organisation handling its own data suffers a breach it is clear that the organisation will be investigating and managing the incident with its own interests as a priority. It has control of its systems and the data residing therein and can make decisions that protect its interests from a business and liability perspective. If a cloud provider suffers a data breach exposing its customers' data, its interests may not be (and perhaps often are not) aligned with its customers'. To the extent the service provider faces potential liability; its handling of a breach situation may favour its own interests. Cloud customers may not have the control over or access to the system they would typically enjoy in order to investigate, gather evidence and remediate a data breach. Service providers may be encouraged to withhold certain information from their customers to protect themselves. Unfortunately, while data loss and data leakage are both serious threats to cloud computing, the measures put in place to mitigate one of these threats can exacerbate the other. For this reason, we are implementing a three level (authentication, confidentiality and data integrity) security to ensure protection at all levels, in addition to secure disposal of information.

2 RELATED WORK

Cloud Computing is an agile, reliable, cost effective and scalable method for delivery of computing and delivery of data. End users access cloud based applications through a web browser or a lightweight desktop or mobile app while the business software and data are stored on servers at a remote location. Cloud Computing is an emerging computing paradigm in which resources of the computing infrastructure are provided as services over the Internet. To keep sensitive user data confidential against untrusted servers, existing solutions usually apply cryptographic methods by disclosing data decryption keys only to authorized users. A number of researches have been conducted to evaluate the security issues in cloud computing including control measures to reduce threats through various cryptographic algorithms, risk assessments, secure disposal of information and secure storage of data.

Over the years, a lot of researches have been conducted on the risks associated with cloud computing and a number of solutions have been proposed in which most of them are not in use or implemented. V. Gampala, et al, analyzed security challenges in cloud computing with the risks associated with them, and proposed a number of theoretical solutions to them. They explored data security in cloud computing by implementing digital signature and encryption with elliptic curve cryptography to provide confidentiality and authentication of data between clouds without stating ways to apply them [3].

R. Sridevi and V. Bande, considered diverse security sphere parameters in cloud computing such as Framework, Risk Management, Compliance, Lifecycle Management, Interoperability, Business Continuity, Data Center Operations, Incident Response, Encryption and Key Management, Identity and Access Management, Virtualization, Static Access Security, Internet Access Security, and Dynamic Access Security [4]. They used the classification and survey results to discover similarities and explored the differences in the architectural approaches of cloud computing and also to identify areas requiring thorough research. They provided findings based on the detailed review that could assist in analyzing best fit scenario for an elegant secured cloud computing environment. They concluded that, multi- provider cloud environment cannot provide complete security. They further proposed that to offer complete security trust, assessment work would be carried out under the supervision of third party module (TPM).

R. Sunita and G. Ambrish proposed encryption using hybrid algorithm and secured endpoints whereby a Hybrid Algorithm is used to encrypt the message by which firstly the password will be encrypted by the Ceaser cipher then the encrypted result will again be encrypted by using RSA substitution algorithm and finally the result will again be encrypted by the mono alphabetic substation method [5]. Then the password will be sent to the server with the plaintext user name and if it matches only then the user get access to the system.

P. Arora and A. S. Tyagi describe the performance of different security algorithms on a cloud network and also on a single processor for different input sizes and Advanced Encryption Standard security algorithm was implemented for ensuring security framework [6]. They implemented various cryptographic algorithms on a cloud network in which they concluded that the algorithms implemented are more efficient than using them on single system. The simulation was done on the eclipse and the graphical results were shown by using mat lab. They also stated that performance of an algorithm on

a cloud network varies according to the type of the algorithm such as symmetric, asymmetric or hashing and also varies with the size of the input. This was a simulation and not actual implementation so it may not be very accurate.

NIST (the National Institute for Standards and Technology) of the U.S. Department of Commerce derives its mandate from the U.S. Constitution, through the congressional power to "fix the standard of weights and measures." In brief, NIST establishes the basic standards of science and commerce. Whatever NIST says about cryptography becomes implemented in cryptographic applications throughout U.S. government agencies. Its influence leads to the widespread use of its standards in industry and the broad adoption of its standards internationally [1], [7]. Through the Snowden disclosures, the NIST standard for pseudo-random number generation has fallen into disrepute. Random number generators can either be truly random (obtaining their values from randomness in the physical world such as a quantum mechanical process) or pseudo-random (obtaining their values from a deterministic algorithm, yet displaying a semblance of randomness). A subclass of pseudo-random number generators are cryptographically secure, intended for use in cryptographic applications such as key generation, one-way hash functions, signature schemes, private key cryptosystems, and zero knowledge interactive proofs. A. Priyadharshini made a survey on security issues and the countermeasures in cloud computing storage. To keep sensitive user data confidential against untrusted servers, existing solutions usually apply cryptographic methods by disclosing data decryption keys only to authorized users [8]. This work is aimed at eliminating the computation overhead in countering the security issues in cloud storage by using Kerberos authentication mechanism and addressing the need for moving to multi-clouds.

3 PROPOSED WORK

The proposed work includes authenticating the user by providing challenge response and one-time password system. Risk assessment will be done to identify and prevent potential risks, if a risk is determined, the user will be blocked from accessing the system, otherwise, a one-time password will be sent to the user's valid email and phone number to be used to log into the system. Encryption will be done on uploaded files using elliptic curve cryptography and the key will be managed appropriately to prevent unnecessary exposure to unauthorised users. The elliptic curve cryptographic algorithm will be enhanced using a cryptographically secure random number generation to make the number unpredictable. It also includes secure disposal of information and secure transmission between two clouds as shown in Fig 1. The application is created as Software as a Service and deployed on Google App Engine platform.

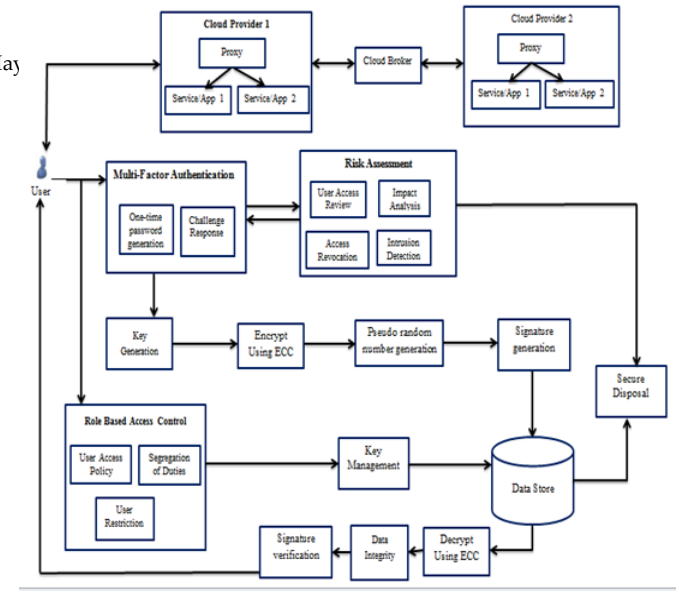

Fig. 1. Security system to prevent data breaching

3.1 Process

When a user requests for access either to upload or download files into the cloud:

1. **Multi-factor authentication:** This is an approach to authentication which requires the presentation of a knowledge factor ("something only the user knows"), a possession factor ("something only the user has"), and / or an inherence factor ("something only the user is"). The system uses challenge response and one-time password schemes for authentication.

Challenge response: This form of authentication is to check validity of the user by providing challenge response module where a series of personal questions will be asked. The challenge response consists of a series of personal questions that the valid user knows of.

One-time password will be sent to the user's mail and phone number for the next login to the system anytime the user logs in to the system. The same password cannot be used twice and it is deleted from the database after the user logs into the system.

2. Risk assessment: This is to prevent unauthorized persons from using either automated attack to obtain passwords or guessing the password because they have unlimited password attempts [4], [9]. Risk is detected if there occur four failed login attempts and the system blocks the user from further attempts by blacklisting the user. A *user access review* is implemented to actively monitor and verify the appropriateness of a users' access to the system and applications. *Impact analysis* is done to evaluate the possible future effects that a particular activity of a user may have on the system or an individual's privacy. *Intrusion detection* mechanism monitors the system activities for malicious activities or policy violations and identifies suspicious patterns that may indicate a system attack from someone attempting to break into or compromise the system. *Access Revocation* is done to blacklist the user if the impact of his activity on the system is high.

3. Encryption: If the user uploads a file, the file will be encrypted using an enhanced elliptic curve cryptographic algorithm and the original file will be deleted in order to avoid it

being intercepted. Elliptic curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. The use of elliptic curves in cryptography was suggested independently by Neal Koblitz and Victor S. Miller in 1985 [5]. Public-key cryptography is based on the intractability of certain mathematical problems. Public-key systems are secure assuming that it is difficult to factor a large integer composed of two or more large prime factors. For elliptic-curve-based protocols, it is assumed that finding the discrete logarithm of a random elliptic curve element with respect to a publicly known base point is infeasible. The size of the elliptic curve determines the difficulty of the problem. The primary benefit promised by ECC is a smaller key size, reducing storage and transmission requirements—i.e., that an elliptic curve group could provide the same level of security afforded by an RSA-based system with a large modulus and correspondingly larger key—e.g., a 256-bit ECC public key should provide comparable security to a 3072-bit RSA public key. For current cryptographic purposes, an elliptic curve is a plane curve which consists of the points satisfying the equation (1):

$$y^2=x^3+ax+b \quad (1)$$

The elliptic curve cryptography involves *key pair generation, signature generation and verification*, and *encryption* and *decryption*. Supposed A wants to send to B an encrypted message. A takes plaintext message M, and encodes it onto a point, PM, from the elliptic group. A will then choose another random integer, k from the interval [1, p-1]. The cipher text is a pair of points shown in (2) which will be sent to B.

$$PC = [(kB), (PM + kPB)] \quad (2)$$

4. Decryption: If the user will download or view a file, the encrypted file will be decrypted using elliptic curve algorithm. As depicted in (3), the receiver of the encrypted file will compute the product of the first point from PC and his private key, then takes the product and subtracts it from the second point from PC then decodes PM to get the message, M.

$$(PM + kPB) – [dB(kB)] = PM + k(dBB) – dB(kB) = PM \quad (3)$$

Cryptographically secure pseudorandom number generation

Choose an arbitrary string A
H= F_H(S, P) // Compute hash
X →H // Convert hash H to field element X
Test x-coordinate for validity on curve E
{ x=true; // If valid, decompress X to obtain point Q
Input P, Q, Seed
Generate random number
Apply second hash, F_H(RN)
New Output}

5. Key and file management: The system will be managed in such a way that even in an organization, there will be multi-level role based security such that every level of the organization's management will have different priorities in accessing to ensure that only authorised user's will be able to access specific information. MD5 algorithm will be used to produce a message digest of the key (also called a hash). The system will send the message digest instead of the key itself, which ensures that no one can eavesdrop and learn keys during transmission.

6. Role Based Access Control: This is an approach to restrict system access to authorized users only. The permissions to perform certain operations are assigned to specific roles and access policies are defined for various users due to their privileges [10]. *User access policy*: Only authorized users are granted access to information systems, and users are limited to specific defined, documented and approved applications and levels of access rights. There is a *segregation of roles* or duties for various levels of authorities of users so that each will be restricted to their own area of access and the users will be restricted on the type of operations they want to execute based on their privileges.

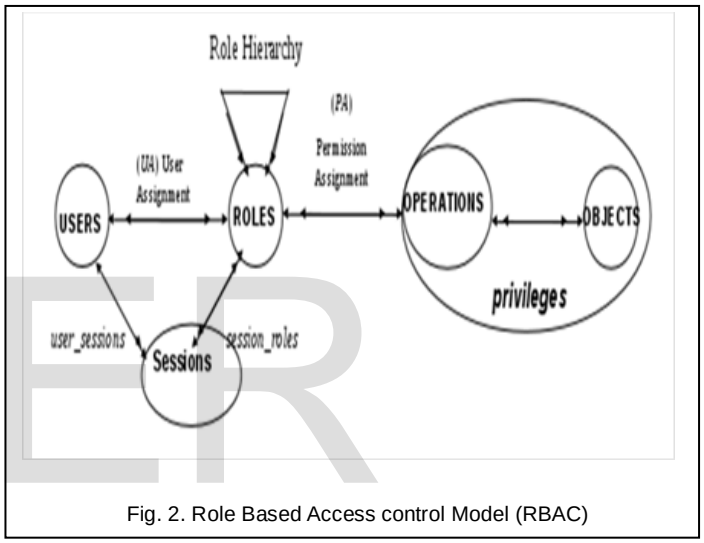

Fig. 2. Role Based Access control Model (RBAC)

Algorithm for role based access control is as follows as depicted in Fig. 2

U, R, P, S, OPS, and OBS (users, roles, permissions, session, operations, and objects, respectively).

i. UA $\subseteq$ U x R, a many-to-many mapping user-to-role assignment relation.
ii. assigned users: (r:R) $\rightarrow 2^U$, the mapping of role r onto a set of users. Formally: assigned_users(r) = {u Є U (u, r) Є UA}.
iii. P = $2^{(OPS \times OBS)}$, the set of permissions.
iv. PA $\subseteq$ P x R, a many-to-many mapping permission-to-role assignment relation.
v. assigned permissions(r: R) $\rightarrow 2^P$, the mapping of role r onto a set of permissions. Formally: assigned permissions(r) = {u Є P (p, r) Є PA}.
vi. Ob(p: P) →{op $\subseteq$ OPS}, the permission-to-operation mapping, which gives the set of operations associated with permission p.
vii. Ob(p: P) →{ob $\subseteq$ OBS}, the permission-to-object mapping, which give the set of objects associated with permission p.
viii. S, the set of sessions.
ix. user sessions (u: U) → 2^S, the mapping of user u onto a set of sessions.

x. session roles (s: S) $\rightarrow 2^R$, the mapping of session s onto a set of roles. Formally: session roles (s_i) $\subseteq$ {r $\in$ R | (session_users (s_i), r $\in$ UA}.

xi. avail_ session_perms (s : S) $\rightarrow 2^P$, the permissions available to a user in a session, $\bigcup_{r \in \text{session\_roles}(s)}$ assigned_permissions (r)

7. Data Integrity: Data Integrity refers to the trustworthiness of system resources, maintaining and assuring the accuracy and consistency of data over its entire life-cycle, and is a critical aspect to the design, implementation and usage of any system which stores, processes or retrieves data. Data integrity can be compromised by multiple factors which include security threads, human errors, physical factors, and software bugs. To ensure integrity of data in the datastore, MD5 algorithm is used. The **MD5** message-digest algorithm is a cryptographic hash function which produces a 128-bit (16-byte) hash value, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity.

8. Secure Disposal: This is the permanent destruction of data that is no longer needed using crypto-shredding, and the use of content discovery to validate that it is not remaining in active storage or archives. Crypto-shredding is the deliberate destruction of all encryption keys for the data; effectively destroying the data until the encryption protocol used is broken or capable of being guessed through brute force mechanism. Crypto-shredding is an effective technique for the cloud, since it ensures that any data in archival storage that's outside physical control is also destroyed once you make the keys unavailable. Shredding is a process of irreversible file destruction, so that its contents could not be recovered that is deleting data without leaving any trace. First, the file is overwritten with byte 9. Next follow one or more passes writing random data. And the last pass overwrites the file contents with byte 246. For two-pass shredding the first pass (which writes byte 9) is omitted. For instance, five-pass shredding is performed as follows: one pass writing byte 9, then three passes writing random data, and then the finishing pass writing byte 246.

9. Secure Transmission between Clouds: Security will be ensured when a user wants to access files in another cloud environment to prevent interception of data in transit, falsification or corruption of data. Kerberos network authentication protocol will be used to protect the system against eavesdropping and replay attacks during transmission at the cloud broker side. The Kerberos protocol is designed to provide reliable authentication over open and insecure networks where communications between the hosts belonging to it may be intercepted. Kerberos is a computer network authentication protocol which works on the basis of 'ticket' to allow nodes communicating over a non-secure network to prove their identity to one another in a secure manner [8]. It is beneficial because it provides mutual authentication both the user and the server verify each other's identity. Because of some limitations of Kerberos it is restricted by some well-known organizations. When the Kerberos server is down due to any physical or environmental attack, no one can log in. This can be resolved by using multiple servers instead of single server. Another limitation with Kerberos is Kerberos assumes that each user is trusted but is using an untrusted host on an untrusted network. Its primary goal is to prevent unencrypted passwords from being sent across that network.

4 Conclusion

Security is a very crucial need in cloud computing so if proper measures are put in place it gives both the service provider and the user a great relief. Our system ensures the corrective measures to protect the integrity of data as well as detecting and preventing possible risks thus ensuring data breaching is prevented.

Our system, however, concentrates on mainly data breaches but there are more threats that cloud security faces. We will therefore in our future scope implement preventive measures for minimizing other threats as well as deploying our system on different cloud platforms to ensure that it can be used across platforms.

References

[1] IEEE CloudCom 2012, 4th IEEE International Conference on Cloud Computing Technology and Science Advances in Computer Science and its Applications , Vol. 1, No. 1, March 2012 , World Science Publisher, United States.

[2] Cloud Security Alliance, The Notorious Nine: Cloud Computing Top Threats in 2013, Internet: http://www.cloudsecurityalliance.org/topthreats, February 2013.

[3] V. Gampala, et al., Data Security in Cloud Computing with Elliptic Curve Cryptography, International Journal of Soft Computing and Engineering (IJSCE), Volume-2, Issue-3, July 2012

[4] R. Sridevi and V. Bande, Comparative Study Of Various Existing Security Scenarios In Cloud Computing Environment , Volume 3, No. 9 Journal of Global Research in Computer Science , September 2012

[5] R. Sunita and G. Ambrish, Cloud Security with Encryption using Hybrid Algorithm and Secured Endpoints, International Journal of Computer Science and Information Technologies, Vol. 3 , 2012, pg. 4302 – 4304

[6] P. Arora and A. S. Tyagi, Evaluation and Comparison of Security Issues on Cloud Computing Environment. IOSR Journal of Computer Engineering (IOSR-JCE), Volume 11, Issue 1, May. - Jun. 2013, PP 39-45.

[7] D. Shumow and N. Ferguson, On the possibility of a back door in the NIST SP800-90 dual ECPRNG, http://rump2007.cr.yp.to/15-shumow.pdf, 2007.

[8] A. Priyadharshini, A Survey On Security Issues And Countermeasures In Cloud Computing Storage And A Tour Towards Multi-Clouds, International Journal of Research in Engineering & Technology (IJRET) Vol. 1, Issue 2, July 2013, 1-10

[9] Daniele Catteddu, Giles Hogben, "Cloud Computing: Benefits, risks and recommendations for information security", November, 2009

[10] S. Ravi, G. Serban, F. F. David, D. K. Richard and C. Ramaswamy, Proposed NIST Standard for Role-Based Access Control, ACM Transactions on Information and System Security, Vol. 4, No. 3, August 2001.