ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 3, March 2019

A Review of Data Classification and Data Security in Cloud Environment

Meenal Jain, Ashok Verma

Research Scholar, Software System, Dept. of Computer Science and Engg. Gyan Ganga Intt. of Tech. and Science Jabalpur (M.P.), India

Professor, Computer Science & Engg. Deptt. of Computer Science and Engg, Gyan Ganga Intt. of Tech. and Science Jabalpur (M.P.), India

ABSTRACT: Cloud computing plays a major role by storing the data and it can be arranged by a third party. The major drawback in the cloud field is privacy and security issues. One of the main issue is the data security and privacy of information stored and processed at the cloud service provider’s systems. Despite of all these services provided by cloud, it lags in the major side of security. Cloud computing significantly plays a role in the aspect of effective resource utilization and service consumption. Irrespective of the type of clouds every service providers concentrates on the data residing in cloud servers. However, users still have major security and privacy concerns about their outsourced data because of possible unauthorized access within the service providers. The existing solutions encrypt all data using the different key size without taking into consideration the confidentiality level of data. In this research, we propose a secure cloud computing model based on data classification. Many cloud models minimizes the overhead and processing time needed to secure data through using different cryptography mechanisms with variable key sizes to provide the appropriate confidentiality level required for the data. The increasing volume of personal and vital data brings up more focus on storing the data securely. Data can include basic user information, important documents, and other user’s related contents for classification of user’s data.

KEYWORDS: Cloud Computing, Cloud Security, Data Classification, Cryptography, Confidentiality

I. INTRODUCTION

Cloud computing is one of the most used technology in today's world. Cloud computing simply describes that it is a hard disk that provided to the user by some user credentials, which the user can access the data and store the data through internet. In the cloud the user can increase the capacity of the storage, so that large amount of data can be stored. Cloud systems consist of three different layers Saas, Paas, Iaas which provides different kinds of services. Nowadays, cloud computing is a growing area that involves wide range of new technologies and applications that touches almost every house residents. Among the associated concepts with this area is the Cloud Computing which is basically allows the users to access and use the cloud services and applications via mobile devices. But and as we know, the Mobile devices have number of challenges that limit their performance, such as battery life time, and lack of computing resources and storage. Another important issue in Cloud and Mobile Cloud Computing is the security of the stored data. Cloud storage services are used widely to store and automatically back up arbitrary data in ways that are considered cost saving, easy to use and accessible. They also facilitate data sharing between users and synchronization of multiple devices. But, there are vital data that is processed and stored in the cloud systems. Losing or exposing these valuable data will have huge bad impact on the data owners being individuals or organizations. And so, there is an increasing demand to protect data over the cloud systems. Users fear from uploading private and confidential files to the online backup due to concerns that the service provider might use them inappropriately. Adding to that, there are concerns about their data being hacked and compromised due to the spread of cloud storage successful attacks. The existing cloud storage frameworks use same key size to encrypt all data without taking in consideration its confidentiality level which might be infeasible. Treating the low and high confidential data by the same way and at the same security level will add unnecessary overhead and increase the processing time.

1.1.1 Cloud Computing Model

There are five major actors [1] in cloud computing based on their participation. Cloud consumer or cloud service consumer (CSC) is the one who gets the service from a cloud provider and pays for the service as per the use. Cloud provider or cloud service provider (CSP) is the one who provides the cloud services to the CSC. Cloud auditor is the one who conducts an independent assessment of cloud services, information system operations, performance and security of the cloud implementations. Cloud broker is the one who interacts between CSP and CSC to make the business happen. Cloud carrier is the one who provides the connectivity and cloud services from CSP to CSC.

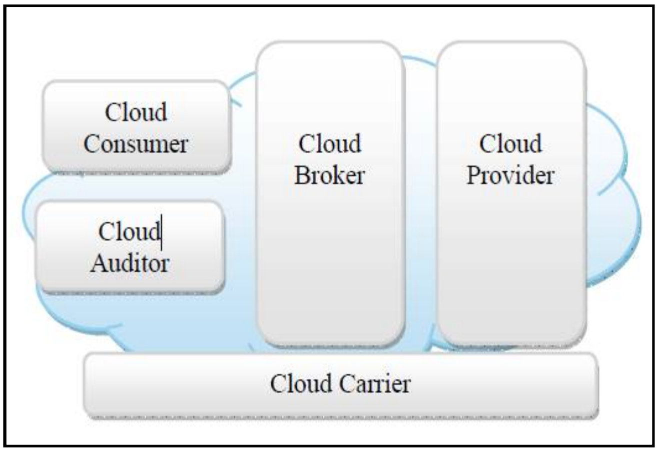

Figure 1.1 Cloud Computing Model

A Cloud can be deployed as Private, Public, Community and Hybrid clouds. Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) are the three service delivery models have become widely recognized and formalized.

II. CLOUD COMPUTING SECURITY

With cloud computing, all your data is stored on the cloud. So cloud users ask some questions like: How secure is the cloud? Can unauthorized users gain access to your confidential data? Cloud computing companies say that data is secure, but it is too early to be completely sure of that. Only time will tell if your data is secure in the cloud. Cloud security concerns arising which both customer data and program are residing in provider premises. While cost and ease of use are two great benefits of cloud computing, there are significant security concerns that need to be addressed when considering moving critical applications and sensitive data to public and shared cloud environments. To address these concerns, the cloud provider must develop sufficient controls to provide the same or a greater level of security than the organization would have if the cloud were not used.

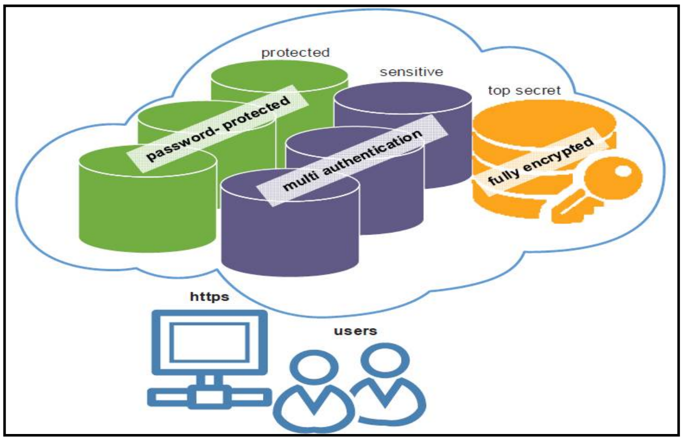

Figure 1.2 Security Classifications to Protect Data in Cloud Storage

. There are three types of data in cloud computing. The first type is a data in transit (transmission data), the second data at rest (storage data), and finally data in processing (processing data). Clouds are massively complex systems can be reduced to simple primitives that are replicated thousands of times and common functional units.

1.2.1 Security Issues in the Cloud Deployment Models

The three deployment models are private cloud, public cloud and hybrid cloud. The security issues of these deployment models are discussed below [6].

A. Security issues in a public cloud

In a public cloud model, the platform and infrastructure are shared among customers. The securities for these services are provided by the cloud service provider. A few of the key security issues in a public cloud include:

1) Since there is no control over the security mechanisms used by the cloud service provider, it is difficult to protect data in all its stages providing the basic requirements of confidentiality, integrity and authenticity.

2) Since most service providers use a multitenant architecture, the possibility of data leakage between the tenants is very high.

3) If the Cloud service provider uses a Third Party vendor for providing the services, then there is added overhead of verifying the agreements and contingency plans between them.
4) There is also a possibility of an insider attack at the service provider side. As the cloud architecture grows the number of insiders grow. Proper laws should be enforced to protect data from malicious insiders.

B. Security issues in a private cloud

A private cloud model enables the customer to have local network and storage space. They provide the flexibility to the customer to implement any kind of required services. There are certain securities issues:
1) Due to virtualization, unauthenticated and unauthorized access to system is possible
2) Malware can be used to attack the host operating system
3) In order to protect from diverse HTTP request the access point of users to access the infrastructure must be protected with standard security techniques.
4) Security policies must be designed to protect attacks from insiders.

The hybrid cloud model is a combination of both public and private cloud and hence the security issues discussed with respect to both are applicable in case of hybrid cloud model. Each of the three ways in which cloud services can be deployed has its own advantages and limitations. And from the security perspective, all the three have got certain areas that need to be addressed with a specific strategy to avoid them [6].

1.2.2 Data security in life cycle

Security content in the cloud resembles traditional security issue and is embodied in all stages of the life cycle. However, as a result of cloud virtualization and multi-tenancy, the content of data security in the cloud has its own unique features. This section analyzes data security problems in the data life cycle. The data lifecycle represents the whole procedure from data production to dumping. The data life cycle is divided into five segments, as shown in Figure.

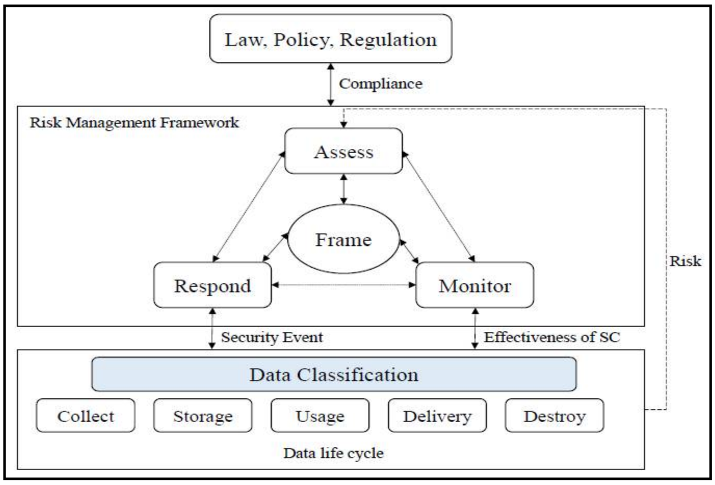

Figure 1.3 Data security in life cycle

III. DATA CLASSIFICATION

Data classification is a very efficient method to protect the data according to its importance and sensitivity. Document NIST Special Publication 800-60 (Guide for Mapping Types of Information and Information Systems to Security Categories) provides a guide for organizations to conduct data classification early in the design of information systems. With the help of the classification, the administrators can identify the application or data to implement corresponding security controls. In this section, we introduce four classification levels (high, medium, and low, normal) according to the different extent of the potential side influence on organizational operations, institutional assets, or individuals.

- Confidentiality: A scarcity of confidentiality is the **unauthorized leak** of data.
- Integrity: A scarcity of integrity is the **unauthorized alteration or damage** of data
- Availability: A scarcity of availability is the **commotion of access** to or use of data.

Classification	Business Information	Personal Information
High Sensitivity	Mid-long term development programs and special plans etc.	Identification card number, passport number etc.
Medium Sensitivity	Important indicator of marketing operations etc.	Email address or contents etc.
Low Sensitity	Company address book etc.	Payment recodes, etc.
Normal Sensitity	Public information (could be found from library or internet)	

Table 1.1 Data Classification

IV. ENCRYPTION ALGORITHMS FOR CLOUD SECURITY

Encryption algorithms have vital role in the field of cloud security. Many algorithms are available for cloud security. Most useful algorithms for cloud security are discussed below.

4.1 Data Encryption Standard (DES)

The Data Encryption Standard (DES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) [15]. It uses single key (secret key) for both encryption and decryption. It operates on 64-bit blocks of data with 56 bits key. The round key size is 48 bits.

Entire plaintext is divided into blocks of 64bit size; last block is padded if necessary. Multiple permutations and substitutions are used throughout in order to increase the difficulty of performing a cryptanalysis on the cipher.DES algorithm consists of two permutations (P-boxes) and sixteen rounds. Entire operation can divided into three phase. First phase is Initial permutation and last phase is the final permutations.

1. Initial permutation rearranges the bits of 64-bit plaintext. It is not using any keys, working in a predefined form.

2. There are 16 fiestel rounds in second phase. Each round uses a different 48-bit round key applies to the plaintext bits to produce a 64-bit output, generated according to a predefined algorithm. The round-key generator generates sixteen 48-bit keys out of a 56-bit cipher key.

3. Finally last phase perform Final permutation, reverse operation of initial permutation and the output is 64-bit cipher text.

4.2 Advanced Encryption Standard (AES)

AES is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST). Most adopted symmetric encryption is AES. It operates computation on bytes rather than bits, treats 128 bits of plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing as a matrix. It operates on entire data block by using substitutions and permutations. The key size used for an AES cipher specifies the number of transformation rounds used in the encryption process [15].

4.3 Rivest-Shamir-Adleman (RSA)

RSA is a public key cipher developed by Ron Rivest, Adi Shamir and Len Adlemen in 1977. It is most popular asymmetric key cryptographic algorithm. This algorithm uses various data block size and various size keys. It has asymmetric keys for both encryption and decryption. It uses two prime numbers to generate the public and private keys. These two different keys are used for encryption and decryption purpose [15]. This algorithm can be broadly classified in to three stages; key generation by using two prime numbers, encryption and decryption. RSA today is used in hundreds of software products and can be used for key exchange, digital signatures, or encryption of small blocks of data [15]. This algorithm is mainly used for secure communication and authentication upon an open communication channel.

While comparing the performance of RSA algorithm with DES and DES, when we use small values of p & q (prime numbers) are selected for the designing of key, then the encryption process becomes too weak and one can be able to decrypt the data by using random probability theory and side channel attacks. On the other hand if large p & q lengths are selected then it consumes more time and the performance gets degraded in comparison with DES. Operation speed of RSA Encryption algorithms is slow compare to symmetric algorithms; moreover it is not secured than DES.

V. OPTIMISING DATA CLASSIFICATION IN DETERMINING SECURITY LEVELS OF PROTECTION

The cloud is a multi-tenant environment, where resources are shared. Threats can happen from anywhere; inside the shared environment or from outside of it. However, placing sensitive data in shared cloud storage is apparently risky. Whether accidental or due to a malicious hacker attack, data privacy, loss or leakage and unavailable for access would be a major security violation involving confidentiality, integrity and availability. It is known that not all data stored in cloud storage are private and confidential. Some are less important and therefore need basic protection. Most CSPs are unwilling to reduce the efficiency of accessing into cloud storages because users expect an equally efficient access into a secured data as the plain text ones. We are emphasizing on protection based on an acknowledged security classification of data determined by the users.

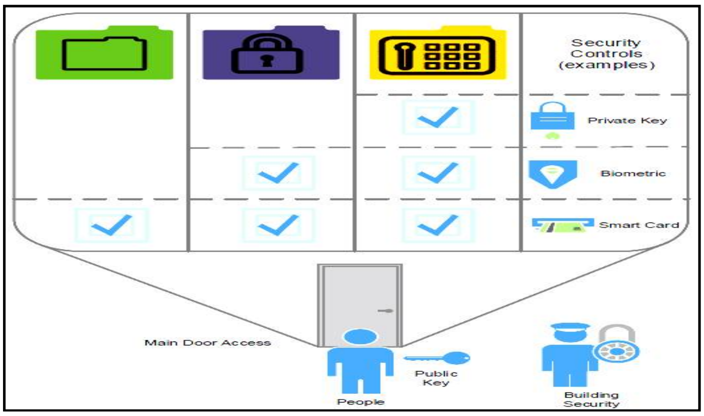

Figure 1.4 Data Protection Levels

There are various ways of protecting a data such as categorizing it into several security groups having different level of protection mechanism as shown in figure. Security classification is known as the process of managing and organising security protection into levels and categories for its most effective and efficient application. A well-planned security classification system makes data protection easier to implement. This can be of particular importance for risk management, legal discovery, and compliance. Assigning a security level of protection to different data classification in cloud storage will give different level of sensitivity to classified information. Users are able to manage their data protection by having assigned a value based on the level of sensitivity. An effective security classification involves a broad awareness of users

Understanding of data residing in cloud storage [8]. Data exists in one of three basic states: at rest, in process, and in transit. All three states require unique security solutions for data protection, but the applied principles of security classification should be the same for each. For an example, data that is classified as sensitive needs to stay sensitive when at rest, in process, and in transit. Data can also be either structured or unstructured. Typical classification processes for structured data found in databases and spreadsheets are less complex and time-consuming to manage than those for unstructured data such as documents, source code, and email. Generally, users will have more unstructured data than structured data. Regardless of whether data is structured or unstructured, it is important for users to manage the level of sensitivity. Therefore when properly implemented, security classification helps ensure that sensitive or classified data are managed with greater oversight than data that are considered public or free to distribute. By having all information on authentication, authorization and encryption will assist user in understanding whether the cloud storage provider supports the data protection requirements mandated by their security classification as below:

A. Authentication

Authentication typically consists of at least two parts: a username or user ID to identify a user and a token, such as a password, to confirm that the username credential is valid. The process does not provide the authenticated user with access to any items or services; it verifies that the user is who they say they are.

B. Authorization

Authorization provides an authenticated user the ability to access an application, data set, data file, or some other object. Assigning authenticated users the rights to use, modify, or delete items that they can access requires attention to data classification. Successful authorization requires implementation of a mechanism to validate individual user's needs to access files and information based on a combination of role, security policy, and risk policy considerations. In ensuring access controls to who can see which and when there must be an effective authenticating system in place.

C. Encryption

Encryption has always been seen as the ultimate security measure to protect data at rest, in process, and in transit. Retaining encryption keys have also been a concern for users storing their data in cloud storage. Further discussion on related work in securing access to cloud storage particularly; authentication, authorization, and encryption are presented in the next section. This will give an overview of previous researches done focusing in this area.

5.1 Data Classification and Data Security in Cloud Computing [14]

Security issues are the greatest threat to any new technology. Cloud computing suffers from many such issues while dealing with securing the data. Confidentiality of data integrity of data as well as availability of data are the key concerns for security in cloud computing. Confidentiality of data is the biggest concern that curbs the quality of services in cloud environment. Two strategies are being used to store the data on cloud storage servers. First is to store the data by encrypting it and second is to directly store the data without encryption. The prime issue of storing the data in cloud is that the user is unaware of where the data is being stored and whether his data is secure or not. So there is always a risk of data confidentiality leakage in distributed computing. Also the data is of different types and have their own different characteristics. So it is very essential to understand the type of data first and then decide which data need to be encrypted and which not. Data classification technique that distinguishes the data into sensitive and non sensitive data. The former is then encrypted via an encryption technique and sent to the cloud. This solves our confidentiality issue in cloud environment.

.

DOI: 10.15680/IJIRCCE.2019. 0703083

VI. CONCLUSION

Both the cloud service provider and the client should make sure that the cloud-data stored is free from all external threats. In order to keep the data secured these threats must be controlled because data residing in the cloud will also cause numerous number of threats and there may arise some security issues, accessibility issues, lack in privacy and reliability of data. Data security in cloud computing is a hard and tiresome task that has not been completely achieved. Various techniques have been proposed for securing data in cloud. Data encryption is a widely used technique for securing the data in cloud. An accurate data security strategy in distributed computing can be decided by first understanding the security necessities of data followed by the selection of possible approach for securing the data. Data security in cloud computing is a hard and tiresome tasking that has not been completely achieved. Various techniques have been proposed for securing data in cloud. Data encryption is a widely used technique for securing the data in cloud. An accurate data security strategy in distributed computing can be decided by first understanding the security necessities of data followed by the selection of possible approach for securing the data. This will help in deciding which data needs to be secured and which not. Data privacy and security are the crucial issues when data stored in the cloud. Data is the valuable asset and of great concerns when moving towards the cloud. Data privacy and security is the active area of research and experimentations in cloud computing. Data leakage and privacy protection is becoming crucial for many organizations moving on to cloud. This research explores the different data security issues in cloud computing in a multi-tenant environment and proposes methods to overcome the security issues.

REFERENCES

[1]. Uma Somani, Kanika Lakhani, Manish Mundra, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing", 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010), ©2010 IEEE.

[2]. Mr. Prashant Rewagad, Ms.Yogita Pawar, "Use of Digital Signature with Diffie Hellman Key Exchange and AES Encryption Algorithm to Enhance Data Security in Cloud Computing", 2013 International Conference on Communication Systems and Network Technologies, © 2013 IEEE.

[3]. Kumar Pal Singh, Dr. Vinay Rishiwal, "Classification of Data to Enhance Data Security in Cloud Computing", © 2018 by IEEE.

[4]. Vanya Diwan, Shubhra Malhotra,Rachna Jain, " Cloud Security Solutions: Comparison among Various Cryptographic Algorithms", © 2014, IJARCSS.

[5]. Gitanjali, Dr. Kamlesh, "Securing Big Data Over Cloud Using Classification and Encryption Techniques", IJRECE VOL. 6 ISSUE 2 APR-JUNE 2018.

[6]. Rizwana Shaikh, Dr. M. Sasikumar, "Data Classification for achieving Security in cloud computing", Procedia Computer Science 45 (2015) 493 – 498, ScienceDirect.

[7]. Lo'ai Tawalbeh , Nour S. Darwazeh, Raad S. Al-Qassas2 and Fahd AlDosari, "A Secure Cloud Computing Model based on Data Classification", Procedia Computer Science 52 (2015) 1153 – 1158, ScienceDirect.

[8]. Amiza Amir, Bala Srinivasan and Asad I Khan," A Communication-Efficient Distributed Algorithm for Large-scale Classification within P2P Networks", SoICT 2015, December 03-04, 2015, Hue City, Viet Nam c 2015 ACM. ISBN 978-1-4503-3843-1/15/12.

[9]. Tina Francis, Dr. Muthiya Madiajagan and Dr. Vijay Kumar, "Privacy Issues and Techniques in E-Health Systems", SIGMIS CPR '15, June 4–6, 2015, Newport Beach, California, USA.

[10]. Spyridoula Lakka, Christos Michalakelis, Teta Stamati and Dimosthenis Anagnostopoulos, "A framework for the classification of Could Computing Business Models", PCI 2015, October 01-03, 2015, Athens, Greece © 2015 ACM. ISBN 978-1-4503-3551-5/15/10.

[11]. Fara Yahya, Robert J Walters, Gary B Wills, "Protecting Data in Personal Cloud Storage with Security Classifications", Science and Information Conference 2015.

[12] Lei Ding, Malek Ben Salem, "A Novel Architecture for Automatic Document Classification for Effective Security in Edge Computing Environments", 2018 Third ACM/IEEE Symposium on Edge Computing.

[13] Miraj Hossain, Md. Rafiqul Islam, "A Model for Ensuring Data Security to Distributed Financial System in Cloud Storage", A Model for Ensuring Data Security to Distributed Financial System in Cloud Storage.

[14] Rasmeet Kour, Suparti Koul, Manpreet kour, "A Classification Based Approach For Data Confidentiality in Cloud Environment", 2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS).