THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Data Encryption in Cloud Storage

Rajesh Dan
Department of Computer Science, St. Xavier's College, Kolkata, India
Md. Saiful Hassan S K.
Department of Computer Science, St. Xavier's College, Kolkata, India
Syed Waquar Ahmed
Department of Computer Science, St. Xavier's College, Kolkata, India
Siladitya Mukherjee
Professor, Department of Computer Science, St. Xavier's College, Kolkata, India

Abstract:
With the exponential growth of internet many enterprises are embracing cloud computing. But very few are concern about user privacy and security. Privacy and security are the key issue in cloud computing. With the growing demand of cloud computing every cloud service providers must ensure user data security and privacy. Encryption is a well- known and effective way to protect user data from unauthorized access. But managing the encryption key and method is also a big challenge for cloud providers. While everyday millions of users and losing their data due to security threats, and companies moving their data to cloud systems, it's more important to secure data in the cloud. This paper analyses the feasibility and method to encrypt the data in cloud.

Keywords: *Encryption, Cloud security, Key management, Data storage, Privacy, Zero-knowledge*

1. Introduction

The idea of an "intergalactic computer network" [1] was introduced in the sixties by J.C.R. Licklider, who was responsible for enabling the development of ARPANET in 1969. His vision was for everyone on the globe to be interconnected and accessing data at any site, irrespective of geographical locations. Margaret Lewis, product marketing director at AMD explain it as "It is a vision that sound a lot like what we are calling cloud computing"[2]. Other experts attribute the cloud concept to computer scientist John McCarthy who proposed the idea of computation being delivered as a public utility, similar to the service bureaus which date back to the sixties. Since then, cloud computing has developed, and the first mile stone was the arrival of Salesforce.com in 1999, which provide the concept of delivering enterprise application via a simple website. Next, Amazon web services in 2002 provide cloud based services including storage, computation and human intelligence through Amazon Mechanical Turk. "Amazon EC2/S3 (Elastic compute cloud/Simple Storage Service) was the first widely accessible cloud computing infrastructure service" said Jeremy Allaire, CEO of Brightcove [2], which provides its SaaS (Software as a Service) online video platform to UK TV stations and newspapers. The next big event was in 2009, as WEB 2.0 hit market, and google and other started to offer browser-based enterprise applications, through services such as Google Apps [2].

At present day, cloud computing is most cost effective, flexible and proven to delivery platform for business or consumer IT services over the internet. As it supports distributer system architecture, multi-domain, multi-user its security risks become more effective. At present a major concern in cloud computing is its privacy and security issues, which is more of concern to the cloud services providers. Applying encryption in cloud storage system is not only the answer, but managing the key is also. The question is where to encrypt data and how. The data can be encrypted by the service provides or by any third party before storing in cloud system, or by the user itself.

An "Encryption in the cloud" [3] survey was done by the Ponemon Institute, which was included more than 4000 IT professionals from seven countries. According to the survey about 38% of the respondents says their organization relay on encryption of the data as it is transferred, typically over the Internet, Another 35% says their organization encrypt the data before it is transmitted to the cloud provider, 27%, says their organization performs encryption within the cloud environment, with 16% of those selectively encrypting at the application layer, and 11% letting the cloud provider encrypt stored data as a service. Similarly in terms of key management 36% relied on their organization, 22% on the cloud service provider and 22% on independent third parties. From the survey it can be seen that organizations takes suitable measure to protect data from attack or breaches. Proving encryption is the most effective security mechanism to protect user data and sensitive information [3].

2. Cloud Computing Framework

The following service models are widely used in cloud computing. Saas (Software as a Service): It is a concept of using software remotely without investing in a full single user software license or hardware required for that particular software. Since saas is based on monthly or yearly cost, it reduces the investment. As the software is in remote location managed by the provider, consumer does not need to install, maintain software.

Paas (Platform as a service): This is a layer of cloud computing that provides developers to develop applications or services in the cloud. No need of the investment on hardware or software required for developing the application. Accessing database, server software, network access makes easy development without building them up from scratch. Developers get flexibility, adaptability, security and an option for wide range development area.

Iaas (Infrastructure as a service): It is one of the three important cloud service models. It provides access to computing resources in a virtualized environment, in other words computing infrastructures. It includes virtual server space, network connections, bandwidth, IP addresses and load balancers. The client is given access to the virtualized components in order to build their own IT platforms.

3. Cloud Deployment Models

- Private Cloud: This is a cloud deployment model where only the client or organization associated with it can operate or access the cloud based environment, giving the organization with greater control and privacy. The private cloud model is closer to the more traditional model of local access networks (LANs). This gives user higher security, privacy and more control, improve reliability.
- Public cloud: public cloud provides services to multiple clients using the same shared infrastructure. It is the most recognizable model of cloud computing. Services are provided in a virtualized environment. This model is extensively used for private individuals, who do not need the level of security of private cloud. This model offers scalability, cost efficiency, reliability.
- Hybrid cloud: This is an integrated cloud service utilizing both private and public cloud to perform distinct functions within the same organization. It can be implemented in various numbers of ways, such as, separate cloud provider team up to provide both private and public services as an integrated service.

4. Data Security Issue in Cloud

There are several issues in user data security and privacy in cloud systems.

Many incidents that are typically called “cloud security issue” are in fact different versions of traditional web-application and data hosting problems. Such as phishing, data loss, week password, compromised host running botnets. Amazon botnet incident in 2009 is one of the incidents. According to cnet news, the cloud-based EC2 (Elastic Compute Cloud) was kept jumping this past week by two incidents: a compromised internal service that triggered a botnet, and a data center power failure in Virginia.

The incidents may be classified into categories.

4.1. Data Loss

It’s a situation when for any reason user lost data. The important and said to be secure data lost due to some unexpected circumstances.

On February, 2011 google mail (gmail) starts deleting user messages [4].

On January, 2009, Magnolia Suffers Major Data Loss. Data gone for good_ [5].

4.2. Hack

Most of the time, hacking results data loss or user’s private and protected data get compromised.

On June 11, 2014, Evernote suffered a crippling distributed Denial-of-service attack that prevented customers from accessing their information [6].

On June, 2011 Hack of web hosting firm Distribute.IT leads to loss of hosted data [7].

On February, 2011 eHarmony User Database was hacked [8].

On March 2, 2013, Evernote revealed that hackers had gained access to their network and been able to access user information, including usernames, email addresses, and hashed passwords. All users were asked to reset their passwords. [9] [10].

On September 2011, Global Sign suspended issuing authentication certificates temporarily after an anonymous hacker compromised their servers. [11].

On January, 2009 hackers break into Google’s Gmail system. Compromising google cloud password system GAIA, potentially expose the user information of millions of Gmail users.

There are other several incidents and other category of incidents like auto fail, outage, internal breaches, and Configuration error. Our concern in this paper is about encryption in cloud, managing the key and encrypts the confidential user data.

5. Background of Our Work

In most of the cloud service provider the system follows is, Security by policy. That literally means that “we have access to your data, but trust us, we will not use your data”, rather than security by design. Security by design may be defines as “We have never any kind of access to your data, without your knowledge we can’t access them”. The later approach seems to be better that the previous one. The major concern in encryption in cloud is to generate and store the encryption key. Keeping the key at the same place just makes the encryption pointless. The key not must be kept at the same place at the cipher text. The most of traditional system expect some like Spideroak [19], who preserve zero-knowledge policy, keeps user credential saved in database. Zero-knowledge police means no knowledge about user password or key to encrypt data. Here user must trust that their password is never saved in the database. In our approach we modified existing algorithm to generate key for the encryption, and the overall encryption process to maximize user data privacy.

6. Approach to Encrypt Data in Cloud

The first step to protecting user privacy is to never store or process user password or key in plaintext in server. So our first step is to locally encrypt user password, which is later used to generate key for encryption, and then send it to the server.
Let the encrypted user password is P1.
P1 is generated from the plaintext password of the user locally, in the browser on client software.
Now the encrypted password P1 is at the server end and that will be used to generate key.

6.1. Algorithm 'Generate Key'

```
Step 1: Generate salt, size=16byte (128 bit). The salt is stored in S1, using Algorithm 'Generate Salt'.
Step 2: let n=10,000 (It is the round number; the greater round no, the greater strength).
Step 3: loop from step 4 to step 11 n times.
Step 4: set K1 to blank.
Step 5: repeat step 6 until P1 and S1 has any bytes left.
Step 6: copy next byte from P1 and S1 to K1.
K1 = K1 append P1[1] append S1[1].
Step 7: copy remaining bytes of P1 or S1, if any.
Step 8: K2 = SHA-512(K1).
Step 8: K3 = copy first 32 bytes (256 bit ) of K2.
Step 9: K4 = AES (pt: K1, Key: k3).
Step 10: set P1 = K4.
Step 11: set S1 = K3.
Step 12: The new key is P1.
```

Algorithm 'Generate Salt' generates a Pseudo Random Number using SHA1PRNG method, using large no of rounds. The salt has to be stored in database, as the salt alone cannot be used to generate key. And the algorithm mixes the salt and the encrypted password by each byte, so cryptanalysis become much harder.

Next step is to encrypt the data itself. The key is generated, and the data should be encrypted in such a way that the persons with direct access to the database and physical data storage device such as hard disk, solid state drive, even cannot generate the plain text.

Our approach is to part the data into several data blocks. The hard disks, that store the data itself, only contain the parts of the data, even not the actual file names. The blocks are all equal in size, so by looking at them will not reveal the actual file name or parts. We have used AES encryption algorithm with 256 bit key, the algorithm to actually implementing the encryption is follows…

6.2. Algorithm 'Encrypt File'

Step 1: Get file (Plain text).
Step 2: Compress the file using lossless compression method, in this case we use LZMA, which is faster and give more compression ratio than other like bzip2.
Step 3: Break the file into several equal size file-parts.
Step 4: For all file parts Fp Do the following:

- Key K1 = GenerateKey : For each part a different key is used.
- Fc1 = AES(Fp,K1) : Using key K1, we apply AES 256bit to get first encrypted file.
- Key K2 = SHA-256(Fp) : It generates the SHA of the encrypted file Fp, and use it as the part of the key of next encryption.
- Key K3 = AES(K1,K2) : Apply AES-256 bit , Key K2 to get new Key K3.
- Fc2 = Blowfish(Fc1,K3) : thus we get the final product Fc2.
- Store only salt,K2 generated in the process.

Step 5: Store the all encrypted parts of equal size.

The above approach helps to encrypt the files not as a whole, but as a list of smaller files. All file parts get encrypted using different key. It is very unlikely to get two same key for two different file part on any two file. As the server only stores the salt, so without knowing the user passphrase anyone cannot decrypt the files. And without the knowledge of the different parts, no one can join the files to get the actual file. As the SHA-256 is different for every files, so it makes the key more random and unique.

7. Conclusion and Future Work

The algorithm works perfectly on any file we have tested on. But also when it comes to share a file among multiple users we need another variation of the algorithm. As the approach works on the passphrase given by user, so sharing is not possible until file is decrypted. But this can be done using key-pair, private and public key approach. The no of rounds the algorithm 'Generate Key' uses, is also much important, but also consume some time. This can be improved by improving the encryption algorithm itself. Our future work would be about sharing files without breaking the encryption or without decrypting the files.

8. References

1. Leiner, Barry M. et al. ""Origins of the Internet" in A Brief History of the Internet version 3.32". The Internet Society. 2003-12-10.
2. http://www.computerweekly.com/feature/A-history-of-cloud-computing
3. Ponemon Institute,Sponsored by Thales e-Security, “Encryption in the Cloud”, July 2012
4. Charlie White. "Google Glitch Disables 150,000 Gmail Accounts". Mashable. 28 February 2011
5. http://www.datacenterknowledge.com/archives/2009/02/19/magnolia-data-is-gone-for-good/
6. King, Leo. "Evernote Pounded By Aggressive Cyber Attack". Forbes, June 11, 2014.
7. http://www.zdnet.com/distribute-it-data-unrecoverable-post-hack-1339317149/
8. http://krebsonsecurity.com/2011/02/eharmony-hacked/
9. "Evernote says security has been breached by hackers". BBC News Online. 2013-03-02.
10. Harrison Weber. "After major data breach, Evernote accelerates plans to implement two-factor authentication". The Next Web. 2013-03-05.
11. BBC News - GlobalSign stops secure certificates after hack claim. Bbc.co.uk (2011-09-07)