International Journal of Computer and Internet Security.
ISSN 0974-2247 Volume 7, Number 1 (2015), pp. 1-5

http://www.irphouse.com

Fog Computing: To Reduce Insider Data Theft Attacks In The Cloud By Using Data Mining And OTP Mechanism

Paritosh Pawar, Manish Patil, Anu Pardeshi, L. Khandekar and Prof. S.S.Vanjire.

Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India

ABSTRACT

Cloud computing is one of today's hottest research areas due to its ability to reduce costs associated with computing while increasing scalability and flexibility for computing services. Ensuring the security of cloud computing is a major factor in the cloud computing environment, as users often store sensitive information with cloud storage providers but these providers may be untrusted. Ensuring the security of cloud computing is a major factor in the cloud computing environment, as users often store sensitive information with cloud storage providers but these providers may be untrusted. Cloud computing promises to significantly change the way we use computers and access and store our personal and business information. With these new computing and communications paradigms arise new data security challenges. Existing data protection mechanisms such as encryption have failed in preventing data theft attacks, especially those perpetrated by an insider to the cloud provider.
In this mechanism we are going to secure our cloud using different techniques like data mining, digital signature(SHA-1) and OTP mechanism.

Keywords- cloud security, data mining, digitalsignature , log formation, OTP.

INTRODUCTION

Fog computing is a term for an alternative to cloud computing that puts some kinds of transactions and resources at the edge of a network, rather than establishing channels for cloud storage and utilization. Proponents of fog computing argue that it can reduce the need for bandwidth by not sending every bit of information over cloud channels, and instead aggregating it at certain access points, such as routers. This allows for a more strategic compilation of data that may not be needed in cloud storage right

away, if at all. By using this kind of distributed strategy, project managers can lower costs and improve efficiencies. cloud computing supports better operational efficiency, but comes with greater risks, perhaps the most serious of which are data theft attacks.

The Twitter incident is one example of a data theft attack from the Cloud. Several Twitter corporate and personal documents were ex-filtrated to technological website TechCrunch and customers' accounts, including the account of U.S. President Barack Obama, were illegally accessed. The attacker used a Twitter administrator's password to gain access to Twitter's corporate documents, hosted on Google's infrastructure as Google Docs. The damage was significant both for Twitter and for its customers. While this particular attack was launched by an outsider, stealing a customer's admin passwords is much easier if perpetrated by a malicious insider. Rocha and Correia outline how easy passwords may be stolen by a malicious insider of the Cloud service provider. The authors also demonstrated how Cloud customers' private keys might be stolen, and how their confidential data might be extracted from a hard disk. After stealing a customer's password and private key, the malicious insider get access to all customer data, while the customer has no means of detecting this unauthorized access.

CURRENT METHODOLOGY

Cloud storage system enables storing of data in the cloud server efficiently and makes the user to work with the data without any trouble of the resources. In the existing system, the data are stored in the cloud using dynamic data operation with computation which makes the user need to make a copy for further updating and verification of the data loss. An efficient distributed storage auditing mechanism is planned which over comes the limitations in handling the data loss. In this paper the partitioning method is proposed for the data storage which avoids the local copy at the user side by using partitioning method. This method ensures high cloud storage integrity, enhanced error localization and easy identification of misbehaving server. To achieve this, remote data integrity checking concept is used to enhance the performance of cloud storage. In nature the data are dynamic in cloud; hence this work aims to store the data in reduced space with less time and computational cost.

Data stored and retrieved in such a way may not be fully trustworthy so here concept of TPA(Third Party Auditor) is used. TPA makes task of client easy by verifying integrity of data stored on behalf of client. In cloud, there is support for data dynamics means clients can insert, delete or can update data so there should be security mechanism which ensure integrity for the same. Here TPA can not only see the data but he can access data or can modify also so there should be some security mechanism against this.

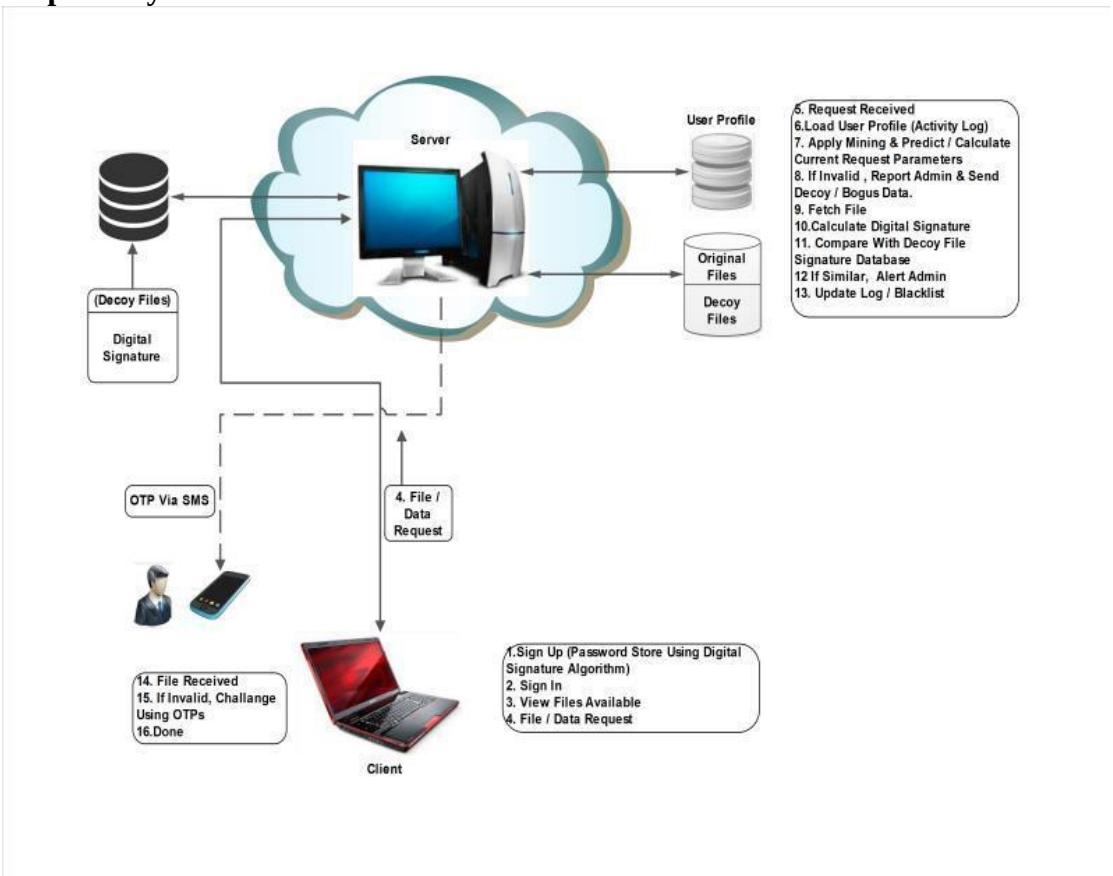
With use of these two we can provide security to cloud but we cannot say system is secure. If the attacker is insider one then we cannot find who is authorise or not. So we provide security from insider attacker with use of our proposed system.

PROPOSED METHODOLOGY

In our proposed system we overcome the previous system's limitations and disadvantages, we are providing security against the insider attacker who may be the authorized person or the outsider who knows the authentication details.

We use concept related to data mining to maintain log tables of each user, by using this we can extract the behavior of every user. Then we use SHA-1 for managing the files. With use of each digital signature we can recognize the modifications in the original file. we use decoy files also, it is used to send the bogus data to the unauthorized users. With use of above security mechanisms if we found user is unauthorized then we are using OTP mechanism for verification.

Proposed system Architecture



1. SHA-1:-

In cryptography, SHA-1 is acryptographic hash function designed by the National Security Agency and published by the NIST as a U.S. Federal Information Processing Standard. SHA stands for "secure hash algorithm".

SHA-1 is the most widely used of the existing SHA hash functions, and is employed in several widely-used security applications and protocols. In 2005, security flaws were identified in SHA-1, namely that a mathematical weakness might exist, indicating that a stronger hash function would be desirable. Although no successful attacks have yet been reported on the SHA-2 variants, they are algorithmically similar

to SHA-1 and so efforts are underway to develop improved alternatives. A new hash standard, SHA-3, is currently under development — an ongoing NIST hash function competition is scheduled to end with the selection of a winning function in 2012.

2. DATA MINING [NAÏVE BAYES]:-

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".

In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

An advantage of the naive Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

3. DECOY FILES:-

The goal is to confuse and confound an adversary requiring more efforts to identify real information from bogus information and provide a means of detecting when an at- tempt to exploit sensitive information has occurred. Decoy Documents" are automatically generated and stored on a _le system by the D3 System with the aim of enticing a malicious user. We introduce and formalize a number of properties of decoys as a guide to design trap-based defenses to increase the likelihood of detecting an insider attack. The decoy documents contain several different types of bogus credentials that when used, trigger an alert. We also embed stealthy beacons" inside the documents that cause a signal to be emitted to a server indicating when and where the particular decoy was opened. We evaluate decoy documents on honeypots penetrated by attackers demonstrating the feasibility of the method.

Decoy information, such as decoy documents, honeyfiles, honeypots, and various other bogus information can be generated on demand and serve as a means of detecting unauthorized access to information and to ‘poison’ the thief’s ex-filtrated information. Serving decoys will confound and confuse an adversary into believing they have ex-filtrated useful information, when they have not.

4. OTP (One time password):-

A one-time password (OTP) is a password that is valid for only one login session or transaction. OTPs avoid a number of shortcomings that are associated with traditional (static) passwords.The most important shortcoming that is addressed by OTPs is that, in contrast to static passwords, they are not vulnerable to replay attacks. This means that a potential intruder who manages to record an OTP that was already used to log

into a service or to conduct a transaction will not be able to abuse it, since it will be no longer valid. On the downside, OTPs are difficult for human beings to memorize. Therefore they require additional technology to work.

OTP generation algorithms typically make use of pseudorandomness or randomness. This is necessary because otherwise it would be easy to predict future OTPs by observing previous ones. Concrete OTP algorithms vary greatly in their details.

CONCLUSION:-

With use of above security concepts we secure our clouds from attackers who are insider to the system. We recognize the user is authenticated or not. We also misguide the attacker with use of decoy files. We also used data mining concept to make log based behavior profiling.

REFERENCES:-

[1] Schneier on Security: Cryptanalysis of SHA-1 (http:/ / www. schneier. com/ blog/ archives/ 2005/ 02/ cryptanalysis_o. html).

[2] Schneier on Security: NIST Hash Workshop Liveblogging (5) (http:/ / www. schneier. com/ blog/ archives/ 2005/ 11/ nist_hash_works_4. html)

[3] Zhang, Harry. "The Optimality of Naive Bayes" (http:/ / www. cs. unb. ca/ profs/ hzhang/ publications/ FLAIRS04ZhangH. pdf). FLAIRS2004 conference.

[4] IEEE- PDDS - Improving Cloud Data Storage Security Using Data Partitioning Technique

[5] Cloud Security Alliance, "Top Threat to Cloud Computing V1.0," March 2010. [Online]. Available:https://cloudsecurityalliance.org/topthreats/csathreats.v1.0.pdf.