INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

INSIDER DATA THEFT DETECTION USING DECOY AND USER BEHAVIOR PROFILE

G. Dileep Kumar[1] and Kolla Morarjee[2]

[1]M.Tech. Scholar, Department of Computer Science and Engineering, CMR Institute of Technology, Medchal, 501401, Hyderabad, Andhra Pradesh, India
dileepkumar.gattu569@gmail.com
[2]Assistant Professor, Department of Computer Science and Engineering, CMR Institute of Technology, Medchal, 501401, Hyderabad, Andhra Pradesh, India
morarjeek@gmail.com

Abstract
Cloud Computing enables multiple users to, share common resources, and to store their personal and business information and access them. The major of the cloud users are from the internet. The users those who have valid authority on the cloud are called insiders. In all the remote users are to be treated as attackers in the security perspective. If the remote user is not an attacker then that should be checked by the security systems. If a valid user's access details are stolen by an attacker, then attacker can enter and access the cloud as a valid user. Distinguishing the valid user and the attacker, the protection of the real user's sensitive data on the cloud from the attacker and securing the fog cloud with decoy information technology are the major challenges in the field of cloud computing. The Decoy Information Technology is used for validating whether data access is authorized; in the eventuality of any abnormal information access detection it confuses the attacker with bogus information.

Keywords: Fog computing, decoy information, decoy technology, data access, and malicious insider.

1. Introduction

Cloud computing is the delivery of computing services over the Internet. The cloud computing has agility, scalability, elasticity and multi-tenancy. Since the sixties, cloud computing has developed. Since the internet only started to offer meaningful bandwidth in the nineties. Now a days, Cloud computing is the delivery of computing services on the Internet. The Cloud services allow individuals and businesses to use software and hardware that are managed by third parties at remote locations. The cloud services include social networking sites, online file storage, online business applications and webmail. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. The cloud computing provides an shared pool of resources, including networks, data storage space, computer processing power, user applications and specialized corporate.

The characteristics of cloud computing include on-demand broad network access, self-service, measured service, resource pooling and rapid elasticity. Self-service means that customers can manage and request their own resources. In the Internet or private networks the services to be offered are known as Broad network access. Pooled

resources, the customers draw from a pool of computing resources, usually in remote data centers. The services can scale larger or smaller, and customers are billed accordingly to the use of a service is measured.

The cloud computing service models are discussed below as:

1) Software as a Service (SaaS): In a Software as a Service model, a pre-made application, along with any operating system, required software, network and hardware are provided. There is no requirement of purchasing a software license, and the vendors run the software application for you.

2) Platform-as-a-service (PaaS): The vendor provides and manages the database, operating system, and everything else needed to run on certain platforms and the customer installs or develops his own software and applications.

3) Infrastructure as a Service (IaaS): The customer installs or develops its own operating systems, software and applications. In this rather than purchasing data center space, software, servers, and network equipment, for these services the vendor provides and bills to clients for the amount of resources consumed.

Cloud services are typically made available via a community cloud, private cloud, hybrid cloud or public cloud. Generally speaking, services providing by a public cloud will be offered over the internet and are operated and owned by a cloud service provider. Some examples include services at the general public, such as e-mail services, online photo storage services, or social networking sites in the web. In public cloud, services for enterprises can also be offered. However, cloud infrastructure is operated solely for a specific organization or a third party. In a cloud community, several organizations share the service and made available only to those groups. The cloud service provider may be owned and operated the infrastructure.

2. Literature survey

Van Dijk et al in [1] proposed Cloud-Application Class Hierarchy that shift towards thin clients and centralized provision of computing resources in the era of cloud computing. It is also strongly illuminated that due to lack of direct resource control there is data privacy violations, abuse or leakage of sensitive information by service providers. The most powerful tool of cryptography i.e. Fully Homomorphic Encryption (FHE) is one the promising tool to ensure data security. The cryptography alone can't enforce the privacy demanded by common cloud computing services by defining a hierarchy of natural classes of private cloud applications and no cryptographic protocol can implement those classes where data is shared among clients. The disadvantage is Abuse and Nefarious use of cloud computing

Iglesias p et al in [2] proposed an adaptive approach is used for creating behavior profiles and recognizing computer users. It presents an evolving method for updating and evolving user profiles and classifying an observed user. As behavior of the user to develop with time, the method is described by fuzzy rules to make them dynamic. It makes use of Evolving- Profile-Library. As a user behavior changes and evolves the classifier is able to keep up to date the created profiles using an Evolving systems approach. It is a one pass, non-interative recursive and can be used in interactive mode. It is operating very efficient and fast as its structure is interpretable and simple. EVABCD can perform almost as well as other offline classifiers in an online environment in terms of correct classification on validation data, and that it can adapt extremely quickly to new data and can cope with huge amounts of data in a real environment with rapid changes. The disadvantage is Insecure Interfaces and APIs.

Rocha F et al in [4] proposed that a malicious insider can steal any confidential data of the cloud user in spite of provider taking precaution steps like. 1) Not to allow physical access. 2) Zero tolerance policy for insiders that access the data storage. 3) Logging all accesses to the services and later use for internal audits to find the malicious insider. It proposes to show four attacks that a malicious insider could do to:- (i) Compromise passwords. (ii)Cryptographic keys. (iii) Files and other confidential data like, clear text passwords in memory snapshots, obtaining private keys using memory snapshots, extracting confidential data from the hard disk and Virtual machine relocation. The disadvantage is Malicious Insiders

Salem B et al in [6] proposed an masquerade for the detection trap-based mechanisms and attacks pose a grave security problem and detecting masqueraders is very hard. The use of trap-based mechanisms as a means for detecting insider attacks is used in general. The use of such trap-based mechanisms for the detection of masquerade

attacks. The desirable properties of decoys deployed within a user's file space for detection. The trade-offs between these properties through two user studies, and proposes recommendations for effective masquerade detection using decoy documents based on findings from the user studies. The different deployment-related properties of decoy documents and a guide to the deployment of decoy documents for effective masquerade detection. The disadvantage is Shared Technology Issues and Data loss or leakage.

3. Proposed Method

The numbers of cloud-based services describe methods to store files, media, and documents, in a remote service that may be accessed wherever a user may connect to the Internet. In suffering problem before such services are broadly accepted and guarantees for securing a user's data, it guarantees only the user and not for others to gain access of data. The main security problem is to provide confidential information, to date, has not providing the levels of assuring. The numbers of proposals have been made to secure remote data in the Cloud using standard access controls and encryption. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including misconfigured services, insider attacks, buggy code, creative construction and the faulty implementations of effective and sophisticated attacks not imagine by the implementers of security procedures [3]. Building a trustworthy cloud computing environment is not good, because continuous accidents are happening, and when they do, and there is no way to get it back, information gets lost. One needs to prepare for such accidents.

The basic idea is that we can limit the damage of stolen data if we decrease the value of that stolen information to the attacker. We can achieve this through a 'preventive' disinformation attack. We suggest that secure Cloud services can be implemented given two additional security features:

1) User Behavior Profiling: The access to a user's information in the Cloud will present a normal access. User profiling is a well-known technique that can be applied here to model, how and when, and in the Cloud it finds how much a user accesses their information. Such 'normal user' behavior can be continuously checked to determine whether abnormal access to a user's information is occurred or not. The commonly used method of behavior-based security is used to find fraud detection. These simple user specific features can serve to detect abnormal Cloud access of data transferred [5].

2) Decoys: Decoy information, such as decoy honeyfiles, documents, honeypots, and other bogus information can be generated on demand and serve as a means of detecting unauthorized access to information. Serving decoys will confound and confuse an adversary into believing in useful information have been ex-filtrated. This technology may be integrated with user behavior profiling technology to secure a user's information in the Cloud. Whenever abnormal user access to a cloud service is noticed, the cloud may return the decoy information and delivered in completely legitimate and normal. The true user, who is the owner of the information, identify the decoy information is being returned by the Cloud. In the case of unauthorized access is correctly identified as an, the Cloud security system will deliver bogus information in unbounded amounts to the adversary, thus securing the user's true data from unauthorized disclosure.

The decoys, then, serve two purposes: (1) validating whether data access is authorized when abnormal information access is detected, and (2) confusing the attacker with bogus information.

We have applied these concepts to detect illegal data access to data stored on a local file system by masqueraders. One may consider illegal access to Cloud data by a rogue insider as the malicious act of a masquerader. The experimental results in a local file system shows combining both techniques can show better detection results, as the Cloud used as transparent to the user as a local file system.

A. Combining User Behavior Profiling and Decoy Technology for Masquerade Detection

1) User Behavior Profiling: illegal users of a computer system are familiar with the files on that system and where they are located. Any search for specific files is like to be targeted and limited. A masquerader, who gets access to the victim's system illeglly, is unlikely to be familiar with the contents and structure of the file system. Their search is likely to be occurring and untargeted. Here we profiled user search developed and behavior user models trained with one-class support vector machines. The one-class modeling stems important for the ability of building a classifier without having to share data from different users. The privacy of the user and their data is to be protected.

2) Decoy Technology: We placed traps within the file system. The traps are decoy files downloaded from a automated service, an Fog computing site that offers several types of decoy documents such as medical records, tax return forms, credit card statements, etc. [13]. The decoy files are downloaded by the legal user and placed in highly-obvious locations that are not likely to cause any interference with the normal user activities on the system. The decoy documents carry a keyed-Hash Message Authentication Code (HMAC), which is hidden in the header section of the document. The HMAC is computed over the file's contents using a key unique to each user. When a decoy document is loaded into memory, we verify whether the document is a decoy document by computing a HMAC based on all the contents of that document. The advantages of placing decoys in a file system are threefold: (1) the activity to detecting of masquerade (2) the confusion of the attacker and the shows the bogus data instead of real information and it additional costs.

3) Combining the Two Techniques: The correlation of search behavior difficult to classify with trap-based decoy files should provide stronger evidence, and therefore improve a detector's accuracy. We assume that detecting of abnormal search operations performed in order to an illegal user opening a decoy file will confirm the suspicion that the user is indeed especially fraudulently another victim user. This method covers the threat model of illegal access to data in the Cloud. Detecting abnormal search and decoy traps together may make a very effective masquerade detection process. Using this two techniques improves detection accuracy.

We use decoys as an oracle for validating the alerts issued by the sensor monitoring the user's file search and access behavior. Instead, we made sure that the decoys were conspicuous enough for the attacker to access them if they were indeed trying to steal information by placing them in highly conspicuous directories and by giving them enticing names.

4. Cloud insider attack detection

The Fog Computing Validation requires

System 1: Test Web Application

1. The application should be deployed on a cloud server (VMware ESX Server). 2. The Application is used to test and to validate the Fog Computing System Detection. The Test Web Applications are the basic inputs for Fog Computing. All the applications should provide the following options It should store user name, password, confirm password and at least ten secrete questions at the time of account creation It should allow forgot password option by querying the user with randomly selected secret questions.

System 2: Fog Computing System

1. To profile or store the user access behavior 2. It analyzes the present behavior with the past profile. The system has to process the system is an interface to view the anomalous user accesses. It allows the administrators to enforce allow/reject policies for the remote users. It provides logs of anomaly detection system 1. User Access Behavior Profiling 2. Decoy File System Maintenance 3. Anomaly Detection 4. Challenge Requests.

User Access Behavior Profiling

The module is concerned about storing the user's request to files on the web application. The module records how many files read and how often. The operations include create, read, write, delete Fig. 3.

Decoy File System Maintenance

For each newly created folder or a file, corresponding decoy file will be maintained. The directory and file structure are same for both the decoy file system and the original file system. The information contained in the decoy file is not original.

Anomaly Detection

The current logged in user access behavior is compared with the past behavior of the user. If the user behavior is exceeding the threshold value or a limit, then the remote user is suspected to be anomaly. If the current user behavior is as the past behavior, the user is allowed to operate on the original data.

Challenge Requests

If the current user's behavior seems anomalous, then the user is asked for randomly selected secret questions. If the user fails to provide correct answers for a certain limits or threshold, the user is provided with decoy files. If the user provided correct answers for a limit, the user is treated as normal user.

System 3: Web Server

It provides an environment to deploy the application. On every access, it stores or log the following details

Client IP, Uid, PID, Time Stamp, Request, Response Code, Response Length, Referrer and User-Agent.

System 4: Internet Users
In internet, the user of the cloud from anywhere.

System 5: Administration System
The system is an interface to view the anomalous user accesses. It allows the administrators to enforce allow/reject policies for the remote users. It provides logs of anomaly detection system. The system is an interface to view the anomalous user accesses. It allows the administrators to enforce allow/reject policies for the remote users. It provides logs of anomaly detection system.

5. Conclusion

In this paper, we present an approach to securing personal and business data in the Cloud. We propose monitoring data access patterns by profiling user behavior to determine if and when a malicious insider illegal accesses someone's documents in a Cloud service. Decoy documents stored in the Cloud alongside the user's real data also serve as sensors to detect illegal access. Once unauthorized data access or exposure is suspected, and later verified, with challenge questions for instance. Such preventive attacks that rely on disinformation technology, could provide unprecedented levels of security in the Cloud and in social networks.

REFERENCES

[1] M. Van Dijk and A. Juels, "On the impossibility of cryptography alone for privacy-preserving cloud computing," in Proceedings of the 5th USENIX conference on Hot topics in security, ser. HotSec'10. "Berkeley, CA, USA": "USENIX Association", 2010, pp. 1–8.
[2] J. A. Iglesias, P. Angelov, A. Ledezma, and A. Sanchis, "Creating evolving user behavior profiles automatically," IEEE Trans. on Knowl. and Data Eng., vol. 24, no. 5, pp. 854–867, May 2012.
[3] J. Pepitone, "Dropbox's password nightmare highlights cloud risks," June 2011.
[4] F. Rocha and M. Correia, "Lucy in the sky without diamonds: Stealing confidential data in the cloud," in Proceedings of the 2011 IEEE/IFIP 41st International Conference on Dependable Systems and Networks Workshops, ser. DSNW '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 129–134.
[5] M. B. Salem and S. J. Stolfo, "Modeling user search behavior for masquerade detection," in Proceedings of the 14th international conference on Recent Advances in Intrusion Detection, ser. RAID'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 181–200.
[6] M. B. Salem and S. J. Stolfo , "Decoy document deployment for effective masquerade attack detection," in Proceedings of the 8th international conference on Detection of intrusions and malware, and vulnerability assessment, ser. DIMVA'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 35–54.
[7] S. J. Stolfo, M. B. Salem and A. D. Keromytis, "Fog computing: Mitigating insider data theft attacks in the cloud", in Proceedings of the IEEE Symposium on Security and Privacy workshop, 2012.
[8] A salya and Ravi M, "survey on defense against insider misuse attacks in the cloud", in Proceedings of the international journals of advanced computing , vol.5, no.1, 2013.
[9] D. Godoy and A. Amandi, "User profiling in personal information agents: a survey," Knowl. Eng. Rev., vol. 20, no. 4, pp. 329–361, Dec. 2005.
[10] D. Godoy, "User profiling for web page filtering," IEEE Internet Computing, vol. 9, no. 4, pp. 56–64, Jul. 2005.
[11] Cloud Security Alliance, "Top Threat to Cloud Computing V1.0," March 2010.
[12] P. Allen, "Obama's Twitter password revealed after french hacker arrested for breaking into U.S. president's account," March 2010. [
[13] *M. Ben*-Salem and S. J. Stolfo, "Combining a baiting and a user search profiling techniques for masquerade detection," in Columbia University Computer Science Department, Technical Report # cucs-018-11, 2011.