VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

DATA DRIVEN TECHNIQUES FOR NEUTRALIZING AUTHENTICATION AND INTEGRITY ISSUES IN CLOUD

Sethuraman Srinivas[1], Sreepriya Menon[2] and Kamalanathan Kandasamy[2]
[1]IBM Security Services, California United States of America
[2]Amrita Center for Cybersecurity Systems and Networks Amrita School of Engineering, Amritapuri Amrita Vishwa Vidyapeetham Amrita University, India
Email: kssethur@us.ibm.com

ABSTRACT

The security of cloud data is at risk as it is been stored by third party providers. Data tampering and unauthentic access of sensitive data by insiders and outsiders are the persisting security concerns of a cloud environment. Despite of all these security drawbacks, cloud computing finds place in various businesses and health care industries. The level of security requirement needed while adopting cloud computing into health care is much higher. In this paper, we address the authentication and integrity issues in cloud focusing the health care domain. We have proposed three major techniques: Data coloring based watermarking, fog computing and User behavior profiling that resolves the above-mentioned issues. In the end, we provide the experimental results based on our analysis.

Keywords: cloud computing, health care, data tampering, authentication.

1. INTRODUCTION

The advancement of cloud computing is becoming popular among health care organizations. The Cloud computing services offers significant benefits to the Health care industry such as the large scale storage of personal health information and the multi tenancy property that enables the health care data to be shared among different entities [1]. We have three types of cloud environments: Public, Private and Hybrid clouds. The Public cloud model offers its infrastructure and services available to general public, whereas the Private cloud is a cloud infrastructure operated solely for a single organization. The Hybrid cloud infrastructure is a composition of both private and public clouds to perform distinct functions within the same organization. Public clouds are more economical as there is no need for purchasing physical hardware [2]. But compared to a private or hybrid cloud, this provides less security to user's data as the data is made accessible to the general public. Due to the sensitive nature of the data being stored and accessed, many health care organizations are opting for private cloud over public cloud services. Using keystroke dynamics cloud based authentication is explored in [26], and data security in cloud is explored using Anonyumus ID [27].

2. ISSUES IN CLOUD COMPUTING

One of the biggest drawback with Cloud computing is the security risk over its sensitive data. Data breach or data loss by unauthentic users, Insider threats from within the organization , Data tampering by masqueraders, Denial of service attack by cloud providers are the few top security concerns that persist in a cloud environment [3].

3. HEALTH CARE DATA BREACHES

Cloud computing is emerging as a vital technology in health care industry. The medical information of patients needs to be widely shared across various health care professionals to improve diagnosis and public health [4]. In the earlier times patient's medical information were recorded in papers. Maintenance of such paper-based medical records is extremely tedious. Now it has been transformed into digital formats (Electronic Health Records). This provides better access to medical records and keeps patients safer by increasing response and treatment time. Many health care industries use cloud computing technology for exchanging health information.

Breaches and identity theft involving medical data are on the rise. Medical data is becoming the new target of attack by hackers [5]. The personal data must be obtained only for a specified purpose, and must not be disclosed to any third party. The consequences of such health care breaches are very crucial. Such type of attacks can lead to life-threatening changes on medical records posing a great threat to your medical care. Some of the health-care data breach incidents that occurred recently is as follows [6]:

1. A former employee of Northgate Medical in Springfield, Mass., took patient information when leaving the practice. The employee took the health information for marketing purposes. The stolen information included patient names, addresses, phone numbers and birth dates.

and extracted cloud drop values confirms the presence of data tampering by unauthentic users [10].

The need for database security can be aided by the data coloring based watermarking technique in relational databases. Here the watermarking is done only on the sensitive numerical attributes. The data owner embeds the cloud drop value generated by him in each row of the database attribute. The user will extract the watermark from the data before its actual use. The cloud watermark detection algorithm need the original RDB participating, and has not implemented blind detection.

The occurrence of data tampering on relational databases is verified by checking for any deviation between embedded and extracted cloud drop value [11].

Forward cloud generator

2. Einstein Health Network, based in Philadelphia, Generate a normally distributed random number Eni' = suffered a possible data breach after a website database was accidentally left accessible on line. The website database contained patient information that had been entered into the health system websites "Request for information form".

In this paper, we have adopted 3 main techniques (1) Data coloring based cloud watermarking technique [7], (2) Fog computing [8] and (3) User behavior profiling for solving authentication and integrity issues at cloud data focusing the health care domain.

4. MAJOR TECHNIQUES

1) Data coloring based watermarking

The relationship between cloud data owners and providers are not trustworthy. Cloud users are worried about the safety of their data from the third party providers. To address the trust issues and integrity compromises in cloud health care, Data coloring based watermarking technique is adopted. The cloud watermarking is based on the cloud model. The Cloud model is a transform model between qualitative concepts and quantitative data. The cloud drop generation depends on three input parameters, the expectation (Ex) value which is decided by the cloud data owner, then the entropy (En) and the hyper entropy (He) values which is the negotiated value between data owner and the cloud provider [7].

This technique comprises of two main approaches: (1) Forward cloud generator and (2) Reverse cloud generator. The forward cloud generator randomly outputs normally distributed cloud drops values based on the three input parameters described above. These cloud drops are treated as the watermarks and remains unique for each user. The three input parameters can be transformed by the Reverse cloud generator to obtain the initial values. This is obtained by passing the generated cloud drop values as the input to backward cloud generator. The Change in the embedded

NORM(En, He^2).
Generate a normally distributed random number xi = NORM(Ex, Eni^2).
Compute $\mu i = \exp[-(xi-Ex)^2/2(Eni)^2]$.
The value of xi with certainty degree of μi is a cloud drop in the domain.
Repeat the above steps, and generate cloud drops.

Reverse cloud generator

Calculate mean X= [redacted] and variance S^2 = [redacted]
Calculate Ex' = X
Compute En' = [redacted]
Compute He' = [redacted]

2) Fog computing

Fog computing is an extension of cloud computing that provides security in the cloud environment. Here a combination of two approaches is deployed (1) user profile analysis and (2) decoy technology which creates an effective masquerade detection system. [12]The search pattern of a masquerader is widespread and not targeted. The above property of an attacker is used as the baseline for detecting the presence of unauthentic users within the system. The CUSUM algorithm is used in which the average fluctuation in the user behavior is calculated when changed. [13] The decoy files contains bogus information about the related data documents. The real user is capable of distinguishing between original and decoy files. Any access on decoy files indicates the presence of unauthorized users with the assumption that the valid users are familiar with the file system.

The decoy records containing the medical information of patients are generated with text obfuscation technique on the original files. These decoy files are placed along with the real files [14]. Monitoring access to decoy files helps in detecting the unauthentic users. The advantage of placing decoy files are: a) detection of unauthorized user's activity and b) confusing the attacker with bogus data. All the documents carry a keyed-Hash Message Authentication Code (HMAC), which is hidden in the the document [15]. The HMAC is computed over the file's contents using a key unique to each user (MD5) [16]. When any document is opened, we verify whether it is a decoy document by comparing the hidden HMAC with a list of HMACs of decoy documents. If the two HMACs match, the document is deemed as a decoy. Even if the attacker got access to the original document, he won't be able to get the correct information because of the coded form of the information [17].

The cumulative summation algorithm is used to identify the deviation from a benchmarked normal behavior pattern. User behavior is identified and monitored using CUSUM algorithm. Implementation of this algorithm calculates average fluctuation efficiently and thus enhances the accuracy to detect the insider data theft attacks. In this paper we are considering the observation, last modification time of decoy files for the CUSUM algorithm. For a malicious user the average

fluctuation of this observation would be greater when compared with that of a normal user. Therefore a combination of the above techniques provides improved security for the sensitive data stored in a file system [18].

Fog computing algorithm

Generate the decoy files from the original files with text obfuscation method
Place the decoy files along with the real files. The access to the decoy is identified with the help of HMAC technique.
If people get access to useful document, the he won't be able to get the correct information because of the coded form of the document
User behavior is monitored and identified with CUSUM algorithm. The average of the observation is recorded and is compared with average of normal observation
If the two consecutive averages give the fluctuation, this indicates abnormal behavior pattern

3) User behavior profiling

This method is considered for detecting intrusion at database level. It consists of two phases (1) Learning phase and (2) Detection phase. In the learning phase, authorized transaction profiles for users are generated from the audit logs and in the detection phase, the behavior of on line transactions are compared with the existing profiles developed. This helps in detecting malicious transactions in the database. If the transaction is authorized then database IDS allows the transaction to commit into the DBMS. The authorized profiles are generated by extracting information from the log file by the help of DBMS auditing mechanisms. If any user wants to execute a new transaction, then he will submit the raw query to feature selector. The Feature selector's extracts the query submitted by the user and store it as on line transaction profile [9].

The newly created profile is compared with the existing authorized transaction profile for the same user. If on line transaction profile matches with authorized transaction profile then the detection engine allows the particular executable transaction to commit into the DBMS. If there is any deviation in the profiles then, the particular transaction is not allowed to commit into the DBMS and current executable transaction is considered as malicious by the detection algorithm. The learning time depends upon the time taken to extract the information from the log file using the DBMS auditing mechanism and storing it into the transaction profiles [19]. We assume that in the learning phase only valid users are allowed to commit their transactions into the database and it was done in the secure environment, then only it can be ensured that the authorized transaction profiles are containing the genuine information. This technique provides additional layer of security in DBMS as only authorized transactions are executed.

User behavior profiling algorithm

Read the entry from audit log file; Sort the entries by user id
For each user id in the file
{
while (current user id has entry in file) do
{
Call SQL query parsing algorithm along with corresponding user id
}
increment the value of user id by one;
}

SQL query parsing algorithm

Fetch the SQL query as string
/*declare 5 variables as strings*/
cmd_type=null targ_list=null relation_list=null qualification=null modified = null
scan the SQL query word by word separated by space; if first word is SELECT then
cmd_type = SELECT;
targ_list = word after cmd_type Go for to find a word FROM;
/*store the word that is after the word FROM in variable relation_list*/
relation_list = word after FROM word; Go for to find a word WHERE; qualification = word after WHERE word;
else if first word is INSERT then cmd_type = INSERT;
go for to find a word INTO;
/* store the word that is after the word INTO in variable targ_list
*/
targ_list = word after INTO word; else if first word is DELETE then cmd_type = DELETE;
go for to find a word FROM;
/*store the word that is after the word FROM in variable relation_list*/
relation_list = word after FROM word; Go for to find a word WHERE; qualification = word after WHERE word;
else if first word is UPDATE then cmd_type = UPDATE;
/*store the word that is after the word UPDATE in variable targ_list*/
targ_list = word after UPDATE word; Go for to find a word SET;
modified = word after SET word; Go for to find a word WHERE;
qualification = word after WHERE word;
end if
Store the variable values into a file for that user id

Malicious transaction detection algorithm

Read the SQL query issued by the current user Call SQL query Parsing Algorithm
Create new profile for the user
While (current user id has entry in the file)
{
compare the variable values for command type, target list, relation list, qualification, modified of new profile with old profile for the current user
}
if command type and target list,
relation list, qualification, modified of both the files are matching for current user id then

```
{
flag = 1; gotoendtran;
}
Else flag = 0;
}
/* increment value of user id*/ user id= user id + 1;
}
endtran:
if (flag = 1) then Valid user;
Else
Malicious user;
```

5. REASONS FOR CHOOSING THE ABOVE TECHNIQUES

Data represent an extremely important asset for any organization. Confidential data such as personal health information must never be disclosed to outside world. One of the most severe threats in the case of cyber-insider attacks is to compromise the integrity of data due by unauthentic users [22]. A malicious attacker who can access the organization databases can tamper the data stored at the database. In order to detect the presence of unauthentic users and integrity issues at cloud data, we have implemented the above three main techniques.

The Data coloring based watermarking helps in detecting tampering by observing change in the watermark. The Fog computing with the combination of decoy technique and user behavior profiling, detects the presence of unauthentic users at the file system. The third technique adopted is the User profiling algorithm. As described above with this technique, intrusion at the database level can be detected. We take the advantage of database audit logs for creating authentic profiles for users [20]. Later in the intrusion detection phase, these profiles are compared before the on line transaction commands are committed into the database.

6. APPLICATION OF DC TOHDB

Data coloring based cloud watermarking technique on Health care database helps in detecting data tampering at the database level. Hackers who can access the sensitive data of patients can alter their health information leading to faulty medical records. With the help of forward cloud generator algorithm, the data owner generates a lot of cloud drops for watermarking the database. These are randomly generated numeric values which are unique for each user. The number of cloud drop generated will be same as the number of rows needed to be watermarked. Only the sensitive data such as the medication id of patients, stored at the database is watermarked. The watermarked database is maintained at the cloud provider's side. When the data has to be used, the cloud drops are extracted from the colored data. The reverse cloud generator is used to extract the cloud drop values. Final color matching comparison will confirm the occurrence of data tampering by unauthentic users.

7. APPLICATION OF FC TOHDB

The application of fog computing method helps in detecting the presence of unauthentic users over health care data. Fog computing is an amalgamation of two approaches (Decoy technique and User behavior profiling). The decoy documents are related documents containing bogus information. Such decoys are trap files placed along with the real files. These files are generated from the original files by text obfuscation method. The file contains the health care information of patients. Each of the file contains a hashed message authentication code (HMAC) which is hidden in the document.

The HMAC of the decoy documents are computed and are stored at the provider side. The cloud provider monitors the access to the decoy files. When a file is opened by a user the HMAC hidden within the document is compared with the HMACs of the decoy documents. If both the HMACsmatches then the document is deemed as the decoy document. Real user of the system is capable of distinguishing between real and the decoy files. In order to confirm that an attacker is accessing the file, we deploy the CUSUM algorithm. With the CUSUM algorithm the average behavior on the last modification time of decoy files is computed. The last modification time of decoy files for a real user is kept as a benchmark for detecting the fluctuation. In the case of an attacker who is accessing the file the average observation on the last modification time would show greater fluctuation from the benchmarked observation.

8. APPLICATION OFUPA TOHDB

The user profiling is the methodological tool that helps in detecting the presence of unauthentic users among the health care database. The profiles of authentic users created in the learning phase are compared with the on line profiles that are developed in the detection phase. The SQL query parsing algorithm works as a feature selector in extracting the SQL commands issued by the user and transaction profiles are created based on these commands from the SQL audit logs [21]. For each user separate profiles are composed based on the data modification command issued by them. The profile developed for patients will be different from the profile created for Doctors, Nurses or other Health care professionals. The Detection algorithm compares the executable transaction profiles with the authorized profiles. If there is any deviation between theses, then the current transaction is considered as malicious transaction by the detection algorithm.

9. CONCLUSIONS

In this paper we have addressed the authentication and integrity issues in cloud focusing the health care domain. We have proposed three major techniques: Data coloring based watermarking, Fog computing and User behavior profiling that resolves the above mentioned issues. Our experimental results prove that the combination of these techniques can achieve good results in solving the authentication and integrity issues. In future we plan to include the encryption techniques too to make the system more secure.

REFERENCES

[1] Cloud standard customer council: Impact of Cloud Computing on Health care http://www.cloud-council.org/deliverables/CSCC-Impact-of-Cloud-Computing-on-Healthcare.pdf.

[2] A Head-to-Head Comparison of Public, Private and Hybrid Cloud Services http://www.technavio.com/blog/a-head-to-head-comparison-of-public-private-and-hybrid-cloud-services.

[3] Louai A. Maghrabi, the Threats of Data Security over the Cloud as Perceived by Experts and University Students, IEEE, 2013.

[4] https://securityintelligence.com/.

[5] https://www.itgct.com/.

[6] http://www.beckershospitalreview.com/.

[7] Kai Hwang, Deyi Li, Trusted Cloud Computing with Secure Resources and Data Coloring. IEEE 2010.

[8] Salvatore J. Stolfo, Malek Ben Salem, Angelos D. Keromytis, Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud. IEEE 2012.

[9] UdaiPratap Rao, Dhiren R. Patel, Design and Implementation of Database Intrusion Detection System for Security in Database, International Journal of Computer Applications (0975-8887); 2011.

[10] Yu-Chao Liu1, Yu-Tao Ma, Hai-Su Zhang, De-Yi Li, Gui-Sheng Chen2, A Method for Trust Management in Cloud Computing: Data Coloring by Cloud Watermarking, International Journal of Automation and Computing. 2011.

[11] Yong ZHANG, XiamuNIU, Dongning ZHAO, A Method of Protecting Relational Databases Copyright with Cloud Watermark, International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:1, No:3, 2007.

[12] Sayali Raje, Namrata Patil, Shital Mundhe, Ritika Mahajan, Cloud Security Using Fog Computing, Proceedings of IRF International Conference, 30th March-2014, Pune, India, ISBN: 978-93-82702-69-6.

[13] ManreetKaur, Monika Bharti, Securing user data on cloud using Fog computing and Decoy technique, International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 10, October 2014 pg. 104-110.

[14] Brian M. Bowen, Shlomo Hershkop, Angelos D. Keromytis, Salvatore J. Stolfo, Baiting Inside Attackers Using Decoy Documents, Department of Computer Science Columbia University.

[15] Supreet, Dr. Vinay Chopra, Data Security Approach Based On Hmac Algorithm For Cloud Environment, Proceedings of International Conference on Networking and Computer Application July 15-16, 2015, ISBN:9788193137314.

[16] HMAC: Keyed-Hashing for Message Authentication https://tools.ietf.org/html/rfc2104.

[17] Ashadeep, Sachin Majithia, Enhancement in Cloud Data Security Using Fog Computing, International Journal of Engineering Research-Online, Vol. 3. Issue.1, 2015.

[18] EBook: Security and Privacy - Silver Linings in theCloud: 25th IFIPTC 11 International Information Security Conference, SEC 2010, Held as Part of WCC 2010, Brisbane, Australia, September 20-23, 2010, Proceedings.

[19] AshishKamra, EvimariaTerzi, Elisa Bertin; Detecting anomalous access patterns in relational databases, The VLDB Journal (2008), Springer.

[20] Agrata Jain, Sneha Saswade, Yogesh Phalke, Chaitanya Gholap, Intrusion Detection and Forensic Analysis on Database using Log Mining Approach, International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014 533 ISSN 2229-5518.

[21] Zhong Y., Qin X., "Database intrusion detection based on user query frequent itemsets mining with on user query frequent itemsets mining with constraints", In Proceeding of the 3rd nternational conference on information security, pp. 224-225, 2004.

[22] Elisa Bertino, Gabriel Ghinita, Towards Mechanisms for Detection and Prevention of Data Exfiltration by Insiders, ASIACCS '11, 2011.

[23] J. Delshi Julie, Watermarking for Copyright Protection of Databases in Cloud Computing, Indian Journal Of Research, Volume: 2, Issue: 1 January 2013.

[24] Younghee Park, Salvatore J. Stolfo, Software Decoys for Insider Threat, ASIACCS '12, May 2-4, 2012, Seoul, Korea., Copyright 2012 ACM 978-1-4503-1303-2/12/05.

[25] Umesh K. Gaikwad, Shirish S. Sane, Effective Classifier for User's Behavioral Profile Classification, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3), 2014, 4541-4545.

[26] Kumar, T. Senthil, Abhijit Suresh, and Aadarsh Karumathil. "Improvised Classification Model for Cloud Based Authentication Using Keystroke Dynamics." Frontier and Innovation in Future Computing and Communications. Springer Netherlands, 2014. 885-893.

[27] Manimaran, A., and K. Somasundaram. "An Efficient Data Security Mechanism in Cloud Computing Using Anonymous ID Algorithm."