IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Obfuscation as a Security Measure in Cloud Computing

Hitendra[1], Dr. Sandhya Tarar[2]

School of Information & Communication Technology, Gautam Buddha University, Greater Noida, UP, India[1,2]

Abstract: Encryption being a powerful tool for data security fails in the case of cloud computing which also being a very beneficial technology has a lot of issues, biggest of which is privacy as information is stored on third party resources which can be a danger to it and thus cannot be stored in its natural form. Cloud providers offer resources which clients can use to process their data but encryption makes this data unusable on the provider's side thus giving the client only a single choice i.e. either security of data or benefits of cloud computing and putting them in a dilemma. This is the main reason clients shy away from using this great technology. In this paper it is discussed how obfuscation which is semi-encrypting of data so that it does not loose its usable form while reducing the privacy risk, is an optimal security measure for cloud computing. Based on a throughreview of various obfuscation techniques, best technique has been chosen and an algorithm for obfuscation has been proposed in this paper which uses the same technique.

Keywords: Obfuscation, Encryption, Cloud, Security, Privacy.

I. INTRODUCTION

Now days encryption can be termed as the key to making the information technology secure, it is the manipulation of data using an algorithm in such a way that if seen by an unauthorized person it will not make sense but if decrypted properly depending upon the type of cryptography algorithm used, will become perfectly useful for the user. Though it is immensely useful it still cannot solve the the privacy issue at the service provider's side posed by the cloud computing as the data needed at the provider's side should be in its unencrypted form so that it can be processed.

Cloud computing works on the concept of distributed computing where the service provider provides with all the resources i.e. hardware, software and storage in the form of Software as a Service(SaaS), Platform as a Service(PaaS) or Infrastructure as a Service(IaaS) which the client can opt for by paying for it and remotely access these resources using any device such as laptop, smart phone or a tablet which need not have a high end specification. It has various benefits like the services are very cheap, can be chosen as per the requirements of the user and can be opted out the same as per need as well, thus attracts a lot of users and has grown significantly over the years but on the other hand owing to its weaknesses the cloud is still not the first preference of many. When a user is using cloud computing services he is required to transfer the data to be processed to the service provider's side, process the data there and then the output is returned. This service provider basically provides a virtual space to the client along with others clients as well which disturbs the clients the most. Encryption can be used to prevent data from falling into the wrong hands while being transferred but once it reaches its destination needs to be converted into its original form essentially being stored in the same storage space which is provided by a third party and can be in use by competitors as well.

This is the place where obfuscation of data can play a major role. Obfuscation is basically masking of data in such way that the data becomes unusable for an attacker or an unauthorized personnel but still does loose its characteristics which can be used to process the data in this form without affecting the results when the data is de-obfuscated into its original form. Obfuscation is a sub set of encryption and can be referred to as semi-encryption. Since obfuscation allows the data to retain its characteristics, it can be highly useful tool in cloud computing security. Owing to this fact various obfuscation techniques were studied and the best technique in relation to cloud computing security was selected and further a basis for further study was proposed in "Obfuscation as a Measure for Cloud Computing -A Review" [1]. In this paper this technique is further explored and an algorithm which focuses on string type of data as obfuscation of integer type of data is less complex using the said technique has been proposed. Figure 1 depicts the flow of steps which form the basis of this paper.

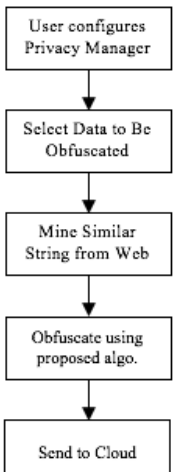

Figure 1: Flow of Steps to Be Followed

DOI 10.17148/IJARCCE.2016.5570

The technique which was selected employs a client side obfuscation tool which obfuscates the data at the client side which may further be encrypted for transfer purpose and then is sent to cloud. This obfuscated data is stored on the cloud and processed in this form only. Once processed the data is sent back and de-obfuscated at the client side only. The client side tool is termed as the privacy manager and also provides additional functionalities such as allowing the user to choose the data or the attributes which need to be obfuscated as not all data is private and obfuscating that data may result in waste of time and resources.

Some scenarios have been taken into account as well as the limitations of the techniques have been studied and reported in this paper as well. Further basis for future studies has been laid and various options which can be explored have been suggested.

II. RELATED WORKS

Following are some studies which have been carried out in need of securing the cloud against privacy and other issues. Their advantages and disadvantages have been discussed which have been used in this paper:

La'Quata Sumter et al. [2] explains that as the cloud computing scope has risen it has also caused an increase of fear because of the security issues in cloud which are increasing unceasingly. According to them sever concerns have cropped up amongst the users relating to the mechanism of access and security in the environment of cloud. A system design has been proposed by them to ensure the security of information of the users that it is safe and cannot be accessed by some personnel who does not have authority to view it. In this System the movement and processing of the information of users is captured which is stored on the cloud. The developed system is case study based and has been tested on a cloud computing environment which is small in size. A claim has been undertaken by them that their model for security is practical in nature as there is a need for a capture device of security on the service provider's side and is apt for cloud computing.

The main beneficial feature of their work is that their model ensures information security for the ultimate user of the cloud. The disadvantage is that the model of security of information on cloud is not appropriate for a scale of cloud computing environments larger than that they have tested it on.

Mowbray & Pearson [3] give a description of privacy manager which resides with the client which reduces the risk and even provides features which are beneficial. A demo has been built by them which proves that the information can be secures by limiting the quantity sensitive information from being sent to the cloud. The benefits that their study is the privacy manager concept which is present at the client side letting him control the information being sent to the cloud. The downside of their research is that the information which is prevented from being sent to the cloud and thus defeating the whole purpose of the cloud computing.

Soren et al [4] describe how embracing of cloud computing has been repressed to such a large extent due to the challenges of privacy, security and safety which actually overshadow the remarkable benefits which are provided by the cloud computing environment. They establish that configuration of services of cloud computing which though a very flexible but are extremely complex and is done using interfaces of web by the user leads to vulnerable threats of security and is the reason that these incidents of security are caused.

The pros of their work is the tool which is proposed by them through which a very thorough analysis of attacks relating to security is provided and vulnerabilities are scanned. This type of analysis helps the providers of cloud computing services to bring improvement in their policies of security and a major drawback is that the work proposed by this study is applicable only to amazon. The contributions of this work would have been more if the it was not limited to a single cloud service.

Pan Yang, XiaolinGui, Feng Tian, Jing Yao &Jiancai Lin [5] in the study of theirs have formulated an algorithm for the obfuscation of data which uses an enhanced method of cloud which generates random numbers having a pseudo nature which in turn is used for the masking of data through obfuscation.
The advantage of the study conducted by them is that it follows obfuscation principles i.e. data being transformed in a form which is unusable to the attacker and yet is retaining its features of characteristics. The main disadvantage of their work is that on numerical data can be protected as it cannot be applied on string type of data.

Table 1: Comparison of Various Obfuscation Techniques and their Pros and Cons

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Techniques	Purpose	Pros	Cons
General	Obfuscation using general and partial masking techniques	Protects data privacy in cloud databases.	Is not successful an AaaS and PaaS cloud architecture
Client-side Obfuscation	The data is obfuscated at the client side and then sent to the cloud	Data privacy from cloud service providers	Operations on a large amount of data needs a lot of time
Encryption	Completely transforming or masking data into another form to ensure data security	Data is most secure when using this technique	Operations on cloud cannot be performed thus defeating the purpose of cloud computing
Noise Obfuscation	Injecting noise along with user queries to confuse cloud service provider	Service provider is unable to distinguish between real and fake requests	Only provides privacy for user activities from monitoring
File split and stored on different clouds	The data is split and then stored on different clouds i.e. storing partial data on a single	Only the user can actually access the whole data	Processing of data may require other parts of the file at one place defeating the purpose

III. OUR SOLUTION

The solution proposed in this paper employees the privacy manager on the client side from literature [2] but does not limit the content that is being sent to the cloud instead the privacy manager deploys the algorithm from literature [5] on numerical data and a hybrid of substitution and shuffling techniques of data masking for string data is used and an algorithm has been proposed for the same. A small application for rating candidates coming for an interview on the salesforce cloud environment was created for the purpose of this study.

A. Client Side Privacy Manager Tool

The privacy manager which is present at the client side has the main feature of obfuscating data but it also provides other services which are discussed here starting with obfuscation:

1) Obfuscation

As stated earlier the main feature of the privacy manager is obfuscation and de-obfuscation of the user's data. This feature employs both the suggested algorithms to carry out obfuscation. The data is obfuscated using a key which is only present with the user which prevents the data from being de-obfuscated without the client's wish. This type of power to control the data become the main appealing feature for the customer.

Nevertheless, there is no need to obfuscate all the data as not all data is private and some data might be needed by the provider for customization and personalization. The time complexity of the obfuscation process depends on the amount of data that needs to be obfuscated i.e. the more amount of data the more will be the time taken to obfuscate it.

2) Preference Settings

This the feature through which the user can customize the preferences for the data. As discussed earlier not all data has a need for obfuscation therefore the user can choose the data which may need it and is actually need by the service provider such as the employee for customized welcome pages for them. This obviously reduces the time of obfuscation process happening at the client side. The rules which are set for an attribute will always apply to the data pertaining to these attributes.

3) Data Access

This is the feature with the help of which violations are detected in the privacy and therefore undertakes the role of mechanism auditor. The privacy manager assists the client to access the information and confirm the exactness, which is theirs, private for them and is being held on the cloud. The feature is activated for private information and is national policy laws based.

4) Feedback

The final feature of the privacy manager is the it allows the customer of cloud services to track their personal information's usage on the cloud. What all information is being transferred such as emails, location etc. can be tracked in the form of feedback by employing the said feature.

B. Algorithm for obfuscating String type of Data

Obfuscation of numerical data can be done using the algorithms from a number of studies such as literature [5] as it is less complex and has a number of techniques.This is a proposed algorithm for obfuscating string type of data which is complex nature. This is a hybrid of substitution and shuffling data masking techniques. The algorithm uses the following:'A' is the array of column string data, 'B' is the array of randomly mined words from the internet with same length as 'A', 'shuffle'is the user/ client defined variable to perform shuffling data mask. See algorithm 3 for obfuscation of String type of data.

Algorithm 1 Obfuscating data OBD (A, B, shuffle)
Input: A, B, C, Words as described above
Output: A' as the obfuscated data
1: n= **ArrayLength**[A];
2: **Set** B[n] = Words;
3: A'[] = B[]; // index of data in A is set equivalent to word
indexed in A'
4: **for** each A'[i] do
5: **swap** A'[i] <-> A'[i+shuffle];
6: **end**

Following screenshots show the obfuscated as well as data in its original form in an app developed for rating interview candidates on the salesforce platform for the purpose of this study:

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

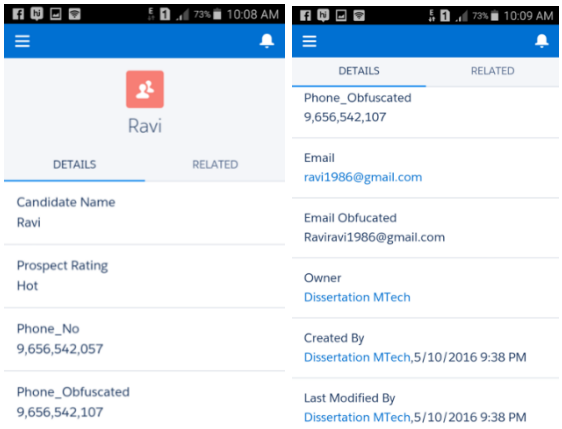

Figure 2: Candidate Information Form

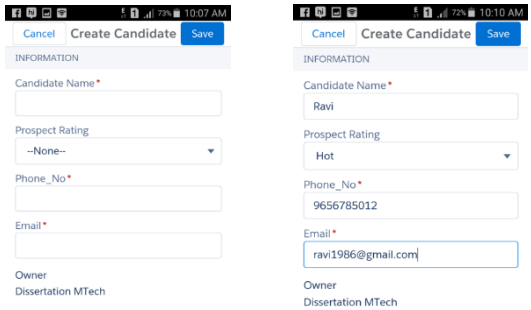

Figure 3: Candidate's Stored Information

IV. WHERE OUR SOLUTION IS APPLICABLE

Many benefits of obfuscation can be yielded to the client especially when the sensitivity of data is very high. Some examples of such type of data are portfolios of security, employee data, formulated strategies of the company etc. Our solution can be applied to such areas which are discussed as follows:

A) Financial data like the quantity of the equity held by a customer, the firms in which their investments are etc. are held in portfolios of security and if gotten into the wrong hands can inflict a lot of harm.

B) Data of employees is another case of highly sensitive data as it holds their names, salary information, address, phone numbers. This information forms the trust of employer – employee relationship n therefore should be taken care of.

C) The strategies which are formulated by the company which maybe for operations, marketing or may be an acquisition is also a critical and extremely sensitive information might need to be distributed to appropriate associates over the cloud and should be done in obfuscated form so that if it falls in the hands of an attacker, it may not yield anything to him.

D) As discussed in the beginning that service providers may also be providing these cloud services to a rival company of its customer. Though the the service provider adheres to the privacy policy but it is a good safety measure to obfuscate any sensitive data sent to the cloud.

V. LIMITATIONS

The user might be able to use these obfuscation techniques to secure data but a reasonable amount of cooperation is required from the service provider as obfuscation and de-obfuscation of the data purely at the client side may require a lot of computing resources at the client side. Another limitation is that it not only depends on the data but also on the service of the cloud and the application which being used by the customer or which requires this data that whether it will be able to work with this obfuscated data or not hence this privacy manger and algorithm may need customization for different and complex applications. For example, there is a cloud application which requires to send customers automated text messages informing them about the status of their request.

In this case obfuscation of data can't be done as the correct mobile phone number is needed at the service providers side or else expensive infrastructure so that text is first forwarded to the obfuscated number along with obfuscated number itself which will be owned by the client of the service provider which has a privacy manager where this number will be de-obfuscated and then forwarded to the real number i.e. the customer of the client of cloud service provider which as can be inferred will be very expensive.
Notwithstanding this, obfuscation is still a very powerful tool and can be used in a number of cloud computing environments effectively providing the customer with the required information security.

VI. CONCLUSION

As a conclusion obfuscation was established as the right technique to go for if the data is highly sensitive and can't be distributed to a third party service and yet the computational power of the cloud is needed for the processing. This study lays down basis for further studies which can be directed towards increasing the efficiency of the privacy manager or developing better algorithms for obfuscation of data made up of string type or both.

VII. ACKNOWLEDGEMENTS

Deepest appreciation is expressed to all those who provided the possibility to complete this research. A special gratitude to GautamBuddha University which gave a platform where this research could be conducted.Furthermore, with much appreciation the crucial role of my mentor at the university, **Dr. Sandhya Tarar** is acknowledged whose contribution in stimulating suggestions and encouragement help me to coordinate the project. I would also like to thank my parents who made me what I am and in the end I would like to thank God who made all things possible.

REFERENCES

[1]. Hitendra, Dr. Sandhya Tarar ― Obfuscation as a Security Measure in Cloud Computing – A Review, IJIACS, Gautam Buddha University, India
[2]. R. La'Quata Sumter, ―Cloud Computing: Security Risk Classification‖, ACMSE 2010, Oxford, USA
[3]. Miranda &Siani, ― A Client-Based Privacy Manager for Cloud Computing‖, COMSWARE'09, 2009, Dublin, Ireland
[4]. Soren Bleikertz et al, ―Security Audits of Multi-tier Virtual Infrastructures in Public Infrastructure Clouds‖, CCSW 2010, Chicago, USA
[5]. Wayne A. Jansen, ― Cloud Hooks: Security and Privacy Issues in Cloud Computing‖, 44th Hawaii International Conference on System Sciesnces 2011
[6]. Pan Yang, XiaolinGui, Feng Tian, Jing Yao, Jiancai Lin - A Privacy-Preserving Data Obfuscation Scheme Used in Data Statistics and Data Mining, Xi'an Jiaotong University, 2013, Xi'an, China
[7]. Flavio Lombardi& Roberto Di Pietro, ―Transparent Security for Cloud‖, SAC'10 March 22-26, 2010, Sierre, Switzerland.
[8]. Mladen A. Vouch, ― Cloud Computing Issues, Research and Implementations‖, Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246
[9]. http://searchcloudcomputing.techtarget.com/definition/Software-as-a-Service
[10]. http://www.v3.co.uk/v3-uk/news/2343547/top-10-cloud-computing-risks-and- concerns/page/5

DOI 10.17148/IJARCCE.2016.5570