Obfuscation Techniques In Cloud Computing: A Systematic Survey

P Nagaraju, Dr N Nagamalleswara Rao,

Abstract: Cloud computing refers to the practice of storing, managing and processing data with the help of a network or remote servers hosted on the internet rather than a local server or a personal computer. Cloud computing offers flexible resources and economies of scale by delivering computing services such as storage, networking, databases, software, servers, analytics and more over the internet which can be said as 'the cloud'. The main reason behind the shifting of traditional way business to cloud computing is with respect to cost, speed, productivity, performance and security. Moreover the security risks should be ensured while sharing the data and resources over the cloud, since there exists security threats such as malicious insiders, hijacking, human error, Distributed Denial of Service attacks and more. Hence, the security and privacy of the cloud should be made more significant by proposing improved techniques on cloud computing. One of the solution is obfuscation which is the most promising technique to prevent and protect the computer systems and networks from above said security threats. In general, obfuscation is the process of converting something difficult to understand using a specifically designed tool called obfuscator which converts the source code automatically in to a program that is much harder to read and understand but works the same way as the source program. This survey aims to enhance the security and privacy of cloud computing with a detailed learning of the obfuscation techniques which rescues the data from malicious attacks in an uncontrolled environment. A systemic review is done on the existing obfuscation techniques and a report is generated that results in the exploration of state of the art in techniques and the algorithms for software obfuscation.

Index terms: Cloud computing, Obfuscation, Security threats, Network, Remote Server

1 INTRODUCTION

Cloud computing is a technology that has been adopted by the recent business environments those who deal with more users and information. The reason why the organizations migrates their application to the cloud is because of the challenges faced in the aspect of providing security [1][2][3] on information, business, physical, operational risk management and more. Through cloud computing the computing resources [4] and other services are shared to the users by the service providers. Three models to offer the cloud services [5] are Infrastructure as a Service (IaaS) [6], Platform as a Service (PaaS) [6] and Software as a Service (SaaS) [6]. The services provided by IaaS includes storage, processing, computing resources and virtual machines, the services provided by PaaS are the computing platforms meant to the operating system and SaaS provides the web browser, a client interface through which the client can access the applications. The rapid development of cloud services [7] means a lot in business, technology and government in one side and the security and privacy [8] is quite challenging in another side. Therefore, security and privacy plays a significant role in protecting the cloud data from the external intruder and also from malicious attacks. In addition, the existing security approaches like cryptographic [9][10][11] and non-cryptographic techniques [12] are too expensive and fails in scalability which in turn made the enterprises and the business environments to lose their trust on cloud computing. Obfuscation [13] is one of the promising techniques that protect the computer systems from harmful malwares. This paper presents a detailed survey on cloud security by obfuscation techniques. In obfuscation, the source program is made harder to read and understand by applying various obfuscation techniques but works the same way as the source program.

Obfuscation is popularly used among game developers, malware and virus developer, industrialists who needs to protect their intellectual property. Obfuscation came in to existence when the cryptographic techniques are not enough to provide security and privacy [8] in protecting the information. Obfuscation is not Encryption – Knowing the difference

Obfuscation is the one where the information is intentionally made difficult to get reversed by a process without knowing the original algorithm that was applied and encryption is a technique where the applied algorithm is well known, reversing the information is done using a secret key. To be simple, obfuscation can be made undo without using any keys while the encrypted data can be decrypted only with the keys. Table 1 depicts the major differences between the obfuscation and the encryption.

Table 1: Major differences between Obfuscation and Encryption

<table>
<thead>
<tr>
<th>Terms</th>
<th>Obfuscation</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is?</td>
<td>Changing the information to a form which is difficult to read and understand, particularly used in context to program codes</td>
</tr>
<tr>
<td>Key?</td>
<td>No key is required</td>
</tr>
<tr>
<td>Data change</td>
<td>Change of data into a difficult form</td>
</tr>
</tbody>
</table>

- Assistant Professor, Department of CSE, GMRIT, Rajam, Srikakulam, AP, India
- Professor, Dept. of IT, RVR & JC College of Engineering, Guntur, AP, India

www.ijstr.org
The remaining portion of this paper contains the following information. Section II reviews the techniques of cloud security rather than obfuscation. A detailed survey on the obfuscation techniques is made in section III and more techniques are discussed and tabulated. Finally in section IV the survey is concluded.

2 Literature review on ‘Not Obfuscation’

This section describes some of the research work apart from obfuscation, means accomplishing the security on cloud through other protection techniques like data masking, public key encryption, conceptual and continuous auditing, encryption techniques, water marking and more. The organizations rely on encryption and other related cloud protection systems only if they deal with extremely sensitive documents and must satisfy the governance obligations. On the other hand, obfuscation is implemented and executed for complicated files and programs and is semantically used in preventing privacy which results in proprietary usage of files and programs. A multilevel classification of various security threats [14] on cloud is discussed in this paper. The evaluation is done by dynamic secureer contract method and the risk levels, type of attacks are identified at each layer of different cloud services and these risks are called low, medium and high as per the ranking. The risk levels intensity is measured based on the cloud layers position and also associated with various cloud services security requirements like data encryption, authentication and authorization, data privacy and more. The infrastructure and the platforms are present in the lower layers and are easily affected by the server attacks. This model leads the cloud service providers to take dynamic decisions in providing the security requirements for the cloud consumer. Also, this model results in providing new dimension to reduce the security threats and minimizes their effects. A public key crypto system called Identity-based encryption (IBE) is proposed [15] which is a revocable scheme consists of a cloud revocation authority (CRA) and improves the performance significantly. The CRA can only hold a secret value that was randomly chosen for all the users and so the security of revocable IBE scheme is retained. In addition, a CRA-aided authentication scheme is constructed with time limit for supporting varied number of cloud services. The experimental results shows that the computation performance is significantly improved and this model is mainly suited for mobile devices and also this model is considered to be semantically secure. Consequently, a detail study of data masking techniques on cloud is proposed [16] that protects the data from various hijackers and thefts. Mainly, different data masking techniques are integrated to secure the data ensuring the trust of the users on the cloud environment. Public cloud, private cloud, hybrid cloud and community cloud are the different cloud development models discussed in this paper. Types of data masking involves Static Data Masking (SDM), Dynamic Data Masking (DDM), shuffling, substitution, random value, number and date variance, nulling out and encryption. Data masking accomplishes increased data theft protection, generates realistic data that improves testing and data sharing. A manual and conceptual continuous auditing (CA) based architecture is proposed [17] to secure clouds. The main contribution of this method depends on third party auditing methodologies for both providers and auditors. The CA methodologies involve Computer Assisted Auditing tools and Technologies, Evidence gathering mechanisms, Auditing system architectures, log inspection and data integrity validation. The architecture of CA includes Data gathering, Data exchange, Data analysis cum presentation and continuous process adjustments. This method assures more reliability and security to cloud service providers. A revocable storage identity based encryption (RS-IBE) is proposed [18] that protects cipher text by providing forward / backward security and also updates the cipher text simultaneously. A concrete model of RS-IBE is constructed to enhance the security level and also to improve both functionality and efficiency resulting in feasible access by both practical application and cost effective data sharing system. Group data sharing technique [19] is proposed enabling some group in the cloud to share data anonymously and storage with high security and efficiency. Tracing the real identities of members on public cloud is happened and an anonymous communication among the group members can be done with respect to the group signature. Cloud computing environment based method called achieving provable data integrity (PDI) is proposed [20] that works well with the client data stored in untrusted servers. For this untrusted storage servers public verification, a simple and efficient audit service is proposed based on bilinear groups that can be called as pairing. Usually, the construction of pairing requires to be built on a Weil or Tate pairing that is normally a super-singular elliptic curve. By exploiting the attribute of bilinear groups, the cost for the auditing protocol initialization and execution is reduced. Furthermore, data dynamics and public verifiability is also supported and this method achieves more efficiency. Cloud computing is made efficient by providing protection to the data according to its classification is proposed [21] in this paper.
This framework comprises the cuckoo algorithm through which the cloud service provider (CSP) can select the appropriate server along with Markov Chain process and Lavy’s flight. This method uses elliptic curve integrated encryption scheme (ECIES) for encrypting the data at the user side and this encrypted data is stored in CSP after the second encryption that was done at the cloud side with advanced encryption standard (AES) technique. This method achieves more efficiency and integrity as it supports integration of server selection approach, authentication and encryption scheme. In addition, confidentiality is also achieved on both user and cloud side through the double encryption process.

3 CLOUD SECURITY AND PRIVACY VIA OBFUSCATION

The process of converting a code semantically to its identical form is termed as obfuscation. This process makes the code very hard to comprehend even when the source code is available with the attackers [22] [23]. The reverse engineering is made harder with applying obfuscation. The possible security threats in cloud computing environment is overcome using this potential obfuscation technique [24] by securing the API’s, the shared applications and infrastructures. This section provides a detailed survey on the currently existing obfuscation techniques which discusses the security overhead, limitations, security in resource sharing and performance efficiency improvements.

3.1 Data Obfuscation with Steganography

To enhance the security of cloud storage, [25] proposed an obfuscation and steganography technique that assures confidentiality. Here, the obfuscated data is embedded inside the image which difficult the process of differentiating the cover image and stego image. The input data is obfuscated using MRADO technique and then the embedding is done using LSB embedding method. Furthermore, the obfuscation techniques are improved by MRADO technique [26] by the process of integrating substitution, transposition and ASCII values. Also, the original value of the information is hidden which cannot be reversed. The process proceeds as follows; from the plain text a list is generated called Line(L) from which the characters(C) are transformed to respective ASCII(ASC) value which is multiplied with the position of the character and results in multiplied output value of each individual character and produces a Numerical Code(NC) to the plain text. Finally, to maintain the line and NC value a look-up table (LT) is created. Modulus operation (MO) is performed on each NC value by 64 resulting in remainder and quotient. Based on Seed(S) (must be a single alphabet or digit), the MRADO square is generated and the obfuscated text(OT) is finally generated with a sequence of Alpha-Digits. This method works well in hiding more data, results in improving the security of data storage.

3.2 Data Obfuscation with Matrices

Similarly, an obfuscation technique without using encryption technique [27] is proposed which preserves the data privacy and confidentiality even in the absence of encryption to cloud servers. Here, the dimension of matrices is resized without using false data. To do so, depending on the size of the matrices, each row and column are divided in to many splits which keeps the number of matrices as it is. Splitting of rows and columns results in varied number of rows and columns respectively. The process involves deceiving the server from knowing the matrices size by exchanging the dimensions of the matrices and hiding or obfuscating the actual values by adding random noise to the matrices value and by shuffling the rows and columns. The limitation here is, the cloud appears more economical.

3.3 Local Differential Privacy Obfuscation

In addition, for IoT data analytics an obfuscation framework is proposed [28] based on edge computing and differential privacy. Here, the privacy is guaranteed with Local Differential Privacy (LDP) data obfuscation. Each server uses LDP framework for protecting the data and then forwards the distilled data to the cloud server. Two steps in LDPO framework are the data distilling using minimax filters and adding randomness [29]. Using minimax filters, the original data is converted to unique bit string by hashing technique. Then the edge servers generate huge number of samples with random noises which introduces obfuscation and the level of privacy is measured through the obfuscation parameter. However, strong privacy is achieved by large obfuscation parameter but the accuracy of the labels is decreased.

3.4 Malware Obfuscation

Malware obfuscation techniques are explored [30], which defeats antivirus scanners. This paper proposes the polymorphic and metamorphic malware based obfuscation technique that includes Dead-code insertion, register reassignment, subroutine reordering, instruction substitution, code transposition and code integration. To change the original appearance of the data, the dead-code insertion adds virtual instructions to a program [31]. To keep the behavior of the program code unchanged, the register reassignment allows switching of registers from generation to generation [32]. Obfuscation on original code by shuffling the subroutines in a random manner is performed in subroutine reordering [33]. The original code is replaced by other equivalent instructions in instruction substitution. Reordering of instructions in original data without changing its behavior is the working of code transposition [34] and the code integration makes the detection and recovery of the original data very difficult.

3.5 Code Obfuscation

Consequently, code obfuscation is proposed [35], which used a code obfuscation engine (CobE) to obfuscate programs by performing code stirring and protect the original data. In code stirring, jump instructions are added to patch the execution path by relocating the small individual code chunks of the binary file. CobE is a hijacking protection mechanism which protects the user applications from reverse engineering attacks. Code stirring on CobE is applied on a binary file with no source code and uses x86-64 system call instructions as jump instructions. Along with code stirring, code injection technique is executed that mix the stirred code chunks generated by the code stirring process with similar code chunks which are not executed results in diverting the attacker. Finally, the GSIM (Generalized Secure in Virtual Machine Monitoring) address location is relocated to distinguish the executed stirred code chunks and the unexecuted code which protects the jump addresses. The limitation is, at some point of the code the obfuscation becomes harder.

3.6 EncryScation (Encryption and Obfuscation)
An EncryScation technique is proposed [36], which is a different method to secure data by applying encryption at the client side and obfuscation at the server side. The obfuscated data is stored on CSP(Cloud Service Provider) database which safely stores both Data Owner/Data User(DO/DU) data and protects from data tampering or misuse. Here, all rights are owned by the owner who manages request to a file or other options. The user can access a file from the authorized owners by locally downloading and decrypting the file and also the user verification of files is done anytime from anywhere. This method protects the user and the service provider at maximum against the hijackers or attackers.

3.7 Code Transformations with Obfuscation
A detailed survey of obfuscation technique is proposed [37] that implements code obfuscation by three different classes such as layout transformation, control-flow transformation and data obfuscation. The alteration of the layout like formatting, comments removal, scrambling and identifiers is performed in layout transformation. In control-flow transformation, the program flow is altered where the computation functionality retains the same. It includes transformations related to rendering, aggregation and redundant computation. The related computations are broken and non-related computations are aggregated in aggregation transformation. Opaque predicates are used in implementation supporting obfuscation to pass through new paths of the control flow graph. The redundant computation includes the dead-code insertion or loop condition extension and finally the data structure of yjr source code is altered in data obfuscation.

3.8 Programming Languages Obfuscation
Obfuscated programming for coding is proposed [38] along with weird languages creation for coding. In literary code evaluation, the programming contexts and the literary contexts are joined using obfuscation and weird languages. In software development, the importance of code reading by human is highlighted by obfuscated code and weird languages. Table 2 depicts some of the other obfuscation techniques in cloud computing domain.

<table>
<thead>
<tr>
<th>Author</th>
<th>Obfuscation technique</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>G. Zhang et al.</td>
<td>Noise Obfuscation</td>
<td>The malicious attacker is made confused by adding a random noise with the customer request.</td>
</tr>
<tr>
<td>Y. Yang et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. Arockiant et al.</td>
<td>Data Obfuscation on client side</td>
<td>A client side key is used to obfuscate the data, which can be de-obfuscated only on the user side but not on the cloud.</td>
</tr>
<tr>
<td>S. Pearson et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K. Kansal et al.</td>
<td>Obfuscation of general data</td>
<td>Obfuscation is done mainly on the identity of the user, data on cloud and the user's behavioral pattern.</td>
</tr>
<tr>
<td>M. Kuzu et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. Bouvry et al.</td>
<td>Obfuscation of source code</td>
<td>This makes the source code harder to read and understand for the attacker so that the cloud software can be protected from reverse engineering.</td>
</tr>
<tr>
<td>R. Omar et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Agir et al.</td>
<td>Obfuscation of location</td>
<td>The user's exact location is made imprecise by doing obfuscation.</td>
</tr>
<tr>
<td>F. Dr et al.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Obfuscation technique</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Celesti et al.</td>
<td>Splittting and storing files on various clouds</td>
<td>To assure data security and availability, obfuscation is done through dividing the data in to various parts and distributed among various cloud service providers.</td>
</tr>
<tr>
<td>M. Villari et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G. Padillii et al.</td>
<td>Obfuscation through encryption</td>
<td>Achieving obfuscation using some cryptographic techniques such as homomorphic encryption, hash function and more.</td>
</tr>
<tr>
<td>G. Gao et al.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Del Ojo et al.</td>
<td>Securing the browser</td>
<td>Obfuscation and hybrid authentication provides high data security and privacy.</td>
</tr>
</tbody>
</table>

4 CONCLUSION
Specifically, this research is focusing on the obfuscation techniques in protecting the information from malicious attacks and from other hijackers. The main target of this proposed research is privacy protection in cloud domain. Improving the security and privacy in cloud computing is the significant role of this study. With the detailed analysis and classification of the existing research methods, we conclude that using obfuscation techniques in cloud computing appears more economical and secure for the user and also provides promising results.

REFERENCES
[10] Ping Li1, Jin Li1, Zhenhang Huang1, Chong-Zhi Gao1, Wen-Bin Chen1, Kai Chen, Privacy-preserving outsourced classification in cloud computing, Cluster Comput (2018) 21, pp 277–286.


[45] B. Bertholon, S. Varrette, and P. Bouvy, Comparison of multi-objective optimization algorithms for the javascript


