International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-10 Issue-1, May 2021

Securing Logs of Functional Testing Infrastructure by Masking Technique

Akshata Khasge, Nagaraja G S

Abstract: ***As organizations become increasingly reliant on technology and data, data protection is now a mission critical priority. Ensuring that data is secure and uncorrupted is essential for avoiding downtime, ensuring productivity, and improving performance. Data protection is keeping the data i.e. confidential data secure from being corrupted or being compromised. Securing the data from logs also is essential when talking about data protection. The testing infrastructure which outputs logs may contain IP address as sensitive data, and this may be misused by other employs of the organization. So, by masking it or removing it from logs will make testing infrastructure more secure and testing can be done without interference of other.***

Keywords: Data Masking, Data protection, Privacy, Security

I. INTRODUCTION

Functional test infrastructure, it is the infrastructure designed to run data domain file system component wise functional test suites. In Functional test infrastructure, details of IPs are required to run the test scripts. The details about client IP are shared once the test scripts are run i.e. IP is visible in console output, in Log files and in yaml file. A log is a record of the events occurring within an organization's systems and networks. These log files contain sensitive data like IP address and password of the client. Infrastructure and system-level administrators need to protect the integrity and availability of log data, and often protect its confidentiality as well. The sensitive data from log is accessible to all the employees of department. To hide this sensitive data from non-admin user, data masking technique is used and is implemented to enhance privacy of sensitive data. Many infrastructures do not consider about the securing of sensitive data. Aim is not to make the whole logs encrypted or inaccessible. There is a risk of tampering the automation scripts by fetching the login credentials from logs.

To prevent the sensitive information from log files being misused by other employees, the methodology used here is Data masking. Data masking is a process of replacing confidential data by using functional fictious data such as character or other data.

Manuscript received on May 17, 2021.
Revised Manuscript received on May 22, 2021.
Manuscript published on May 30, 2021.
* Correspondence Author
Akshata Khasge*, Student, Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru (Karnataka), India. Email: akshatakhasge.scs19@rvce.edu.in
Nagaraja G S, Professor and Associate Dean, Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru (Karnataka), India. Email: nagarajags@rvce.edu.in

II. RELATED WORK

Paper [1] proposes a practical built-in data masking framework (IMETU- Identify, Map, Execute, Test, and Utilize) for BI (Business Intelligence) platform. By keeping in mind about the vulnerability of compromising the health data, this paper highlights the best data masking technique that is best suited to protect privacy of data attributes in health domain and quality analytics of for sensitive data. This paper concentrates on the first two modules of the framework. In Identify module, the PII (personally Identifiable Information) attributes are identified for masking by using library that has masking format. The identifies sensitive attributes gets transformed. Next is Map module which maps the selected personally Identifiable Information data attributes to its respective best fit masking algorithm with the help of an automated system.

Paper [2], is continuation of [1], which proposed the data masking framework IMETU. This paper focuses on the "Execute" and "Test" module of a new proposed data masking technique (COBAD) which is based on the statistical content derived from the loaded dataset at aggregated levels. The next three modules i.e. execute, test, utilize has described and analyzed in this paper. Execute module applies the various masking algorithm in an effective manner. Test module contains testing the results obtained from masking if it has applied successfully. Utilize module involves using reidentification methods to extract the original data to get the right result.

Paper [3], proposes the technique of dynamic data masking, where modified database structure is used which gives true or false credible data. This Data Masking is useful in fields like, protecting sensitive information and in software debugging. The technique used, modifies the masked relation scheme and to make masking more efficient separate masked element is mixed.

Paper [4], proposes a key-based reversible approach of data masking module, which protects data privacy and maintains data utility of the patients record to meet the requirements of data analytics in Business Intelligence platform of the healthcare environment. The module is named as CARE (Create, Apply, Remember, and Employ).

Paper [5], proposes data masking system called ink data masking system. the core technology used here is ink Technology which references to block chain idea.

This technology is based on the private and public key and the digital certificates. It creates the ink code book by using asymmetric encryption.

Paper [6], proposes the new approach of data masking using neural networks. This paper uses associative memory-based approach which falls into the pattern storage classification, so that the stored pattern information can be used in the future to recall. The data set is trained based on the data masking rules. For neural network to learn a rule, input-output pair should be given as input in the binary form.

Paper [7], proposes new approach called multimatrix masking for collusion resistant, which overcomes the security issues of many other masking techniques. Here the data of participants is split into k number of component vectors such that x = v1 + ... + vk, where x is data of participant. Bi is the secret matric shared by all the participants. Then these component vectors are sent to various masking service provider, where these vectors are right multiplied by Bi where the communication is encrypted between each participant and service provider. This system gives k-privacy as there are k masking, this means that privacy of this system is compromised only when all the k masking service providers collude. Once the data is masked, matrix masking is done by sending the masked data to all other masking service provider.

In the above papers, many researchers have proposed the different masking techniques in various other fields and the techniques have applications in various other domains. One interesting fact is that, referred paper have not used one of the masking techniques called “Nulling out or Deletion” , which is the best preferred, when considering security and privacy as an issue

III. KEY TECHNOLOGIES

A. Test Automation

Automation Testing or Test Automation is a software testing methodology that uses special software tools for automated testing to operate a suite of test suits. This can be done by writing scripts for testing or using some testing method for automation. In the System Under Test, automation testing software can also enter test data, compare expected and actual results, and produce comprehensive test reports. Test Automation requires significant cash and resource investment. Successive cycles of production would require repeated execution of the same test suite. It is possible to record this test suite using a test automation tool and re-play it as required. No human intervention is needed once the testing suite is automated.

B. Data Masking

Data masking is an important technique for mitigating the risk of data leakage from inside and outside an enterprise and can be used as a best practice for curing databases that are not productive. To prevent the sensitive information from log files being misused by other employees, the methodology used here is Data masking. Data masking is a process of replacing confidential data by using functional fictious data such as character or other data.

C. Athena Framework

Athena framework is a platform for developing software application. It is created with intension of providing various high quality, reusable and independent components. Athena framework is used so that the libraries written are reused between the test cases.

IV. PROPOSED METHODOLOGY

Functional test infrastructure, it is the infrastructure designed to run ddfs (data domain file system) component wise functional test suites. A log is a record that is generated after the testing, contains details about the IPs, password, cloud provider details, secret key. The sensitive data from log is accessible to all the employees of department. To hide this sensitive data from non-admin user, data masking technique is used and is implemented to enhance privacy of sensitive data. Our aim is not to make the whole logs encrypted or inaccessible. There is a risk of tampering the automation scripts by fetching the login credentials from logs.

Data masking is the method of obscuring-masking, within the data stores, basic data items. It guarantees that sensitive information is replaced by practical, but not actual, data. The aim is that in the designated setting, confidential customer information is not available. Algorithms for masking are designed to be repeatable to preserve referential integrity. Popular business apps require regular patch and update cycles and require 6-8 copies of the application and testing details to be made. Although companies usually have tight controls on production processes, data protection is often left to the employee's trust in non-production situations, with potentially catastrophic consequences. In an automated process, making test and production copies decreases the exposure of sensitive data. The layout of the database frequently changes, so keeping a list of sensitive columns in an application code without rewriting it is helpful.

There are two technique in data masking, suitable for sharing data with unauthorized users Nulling out or deletion and Masking out. First step is to set up a private DNS server. A DNS server is a computer server that contains a database of IP addresses and their associated hostnames. Private DNS is when we use either DNS over TLS (Transport Layer Security) or DNS over HTTPS (Hyper Text Transfer protocol), all the DNS queries are encrypted. Tools used for setting up private DNS server is MobaXterm, PyCharm. Once Private DNS is set up, write a program to map the IPs to respective domain name. Python language is used for programming.

The flow chart of the complete proposed methodology is shown in the figure 1.

Configuration of private DNS Server:

Private DNS is when we use either DNS over TLS (Transport Layer Security) or DNS over HTTPS (Hyper Text Transfer protocol), all the DNS queries are encrypted. BIND is used to configure the private dns server.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication

The prerequisites include, statis ip of Virtual machine. Servers that are in same center, A VPS for primary dns, the statis hostname that has domain.

First step is to install bind and bind-utils packages. First configure the primary dns. BINDs process is called as named. In the named configuration file change the IP to servers static IP and create zones, forward and backward. The forward zone file contains records of mappings between IP and domain name for forward DNS lookups. It resolves domain name to IP. Reverse zone file contains Pointer (PTR) record for reverse DNS lookups. If it receives any query that contains IP, it resolves to FQDN (fully qualified Domain name). Check the bind configuration and start the service. IN most cases, it is good practice to set up secondary dns as well. if primary is unavailable to respond to the requests, the requests goes to secondary dns.

V. IMPLEMENTATION DETAILS

In the testing infrastructure, the log record contains the IP of the VM being used and the DD system on which the testing is carried out. so, these IP can be masked by mapping IPs to their domain.

The input to infrastructure is automated scripts written in python language and the configuration file. The output is the logs that has passed status. The used OS is CentOS7. MobaXterm is used for client connectivity which brings all the unix commands to windows. In testing infrastructure where input is to it is automation scripts, and the configuration file that contains details of the machine on which test is carried out and the individual's workspace IP. In this configuration file instead of giving IP, domain name can be given to mask the IP in logs. When the input is given, the infrastructure scans the Configuration file and finds the domain name of the systems. Using this, it pings the primary server for the IP address of the system.

The primary server searches the records for the mappings of name to IP address. Once it gets the IP address, it returns it to the requested machine. Once the system gets the IP of the DD system, it starts it's testing on that machine. Athena framework is used for this testing and testing is done on DD system. When we use DNS, the IP field is replaced by the domain name, this prevents non-admin users to locate the IP and prevents them from using it and hence builds can run without interference.

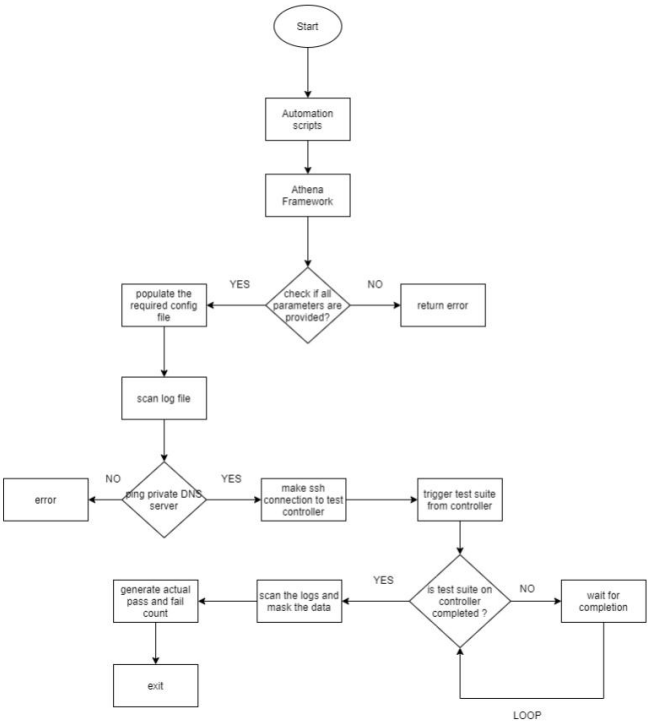

Figure 1. Flow Chart of complete proposed model

VI. EXPERIMENTAL RESULTS

Test cases run on the secure infrastructure shows that the performance of the infrastructure has increased. Hence, the approach of setting up private DNS helps to secure the automation infrastructure resources of the organization and provide stability.

Table-I shows the experimental results of the test cases run on the infrastructure. Before masking efficiency was around 84% and after masking it increased to 90%. With the help of masking technique, the logs are made secure.

Table-I Experimental results

	Total test cases	No. passed	No. failed	Efficiency
Before Masking IP	50	42	8	84%
After Masking IP	50	48	2	>90%

The solution provided, fixes problem of builds failure, scripts tampering in the infrastructure. This solution makes the existing infrastructure more reliable compared to the existing one.

VII. CONCLUSION

This paper has proposed one of the applications of private DNS server in data masking. Few of the testing infrastructure which contain IP in their logs are more susceptible to build failures. The stronger the infrastructure, the more it provides stability, reliability.

Hence, the approach of setting up private DNS helps to secure the automation infrastructure resources of the organization and provide stability. The logs contain the masked IP address, due to which there is less interference during the testing. Hence the builds are made stable. There is greater increase in the performance of the infrastructure and the efficiency of testing the scripts has also increased, making the infrastructure more secure and stable.

REFERENCES

1. Osama Ali and Abdelkader Ouda "*A Classification Module in Data Masking Framework for Business Intelligence Platform in Healthcare*", Electrical and Computer Engineering the University of Western Ontario London, Ontario, Canada, October 2016.
2. Osama Ali (Ozkan) and Abdelkader Ouda, "*A Content-Based Data Masking Technique for A Built-In Framework in Business Intelligence Platform*", IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 978-1-50905538-8, vol. 17, 2017.
3. Zaruhi Aslanyan, Martin Staal Boesgaard, "*Privacy Analysis of Format-Preserving Data-Masking Techniques*", IEEE conference ,978-1-7281-2856,2019.
4. Osama Ali-Ozkan and Abdelkader Ouda, "*Key-based Reversible Data Masking for Business Intelligence Healthcare Analytics Platforms*", IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 978-1-7281-1244, vol. 19, 2019.
5. Fucheng You, Yue Cao, Chenhan Zhang, Chenwei Zhang, "*Data Masking System based on Ink Technology*", 5th International Conference on Information Science and Control Engineering, 978-1-5386-5500-9, vol. 18, 2018.
6. Vishal Anjaiah Gujjary, Ashutosh Saxena, "*A Neural Network Approach for Data Masking*", Article in Neurocomputing, 2010.
7. Samuel S. Wu, Shigang Chen, Abhishek Bhattacharjee, Ying He, "*Collusion Resistant Multi-Matrix Masking for Privacy-Preserving Data Collection*", 2017 IEEE 3rd International Conference on Big Data Security on Cloud, 978-1-5090-6296-6, 2017
8. Osama Ali, Dr. Abdelkader Ouda, "*Secured Data Masking Framework and Technique for Privacy in a Business Intelligence Analytics Platform*", Graduate Program in Electrical and Computer Engineering, 2018.
9. Siddartha B K, Dr. Ravikumar G K , "*Analysis of Masking Techniques to Find out Security and other Efficiency Issues in Healthcare Domain*", Proceedings of the Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (ISMAC) IEEE Xplore Part Number:CFP19OSV-ART; ISBN:978-1-7281-4365-1,660668, 2019
10. Alfredo Cuzzocrea, Hossain Shahriar, "*Data masking Technique for NoSQL Database security*", IEEE International Conference on Big Data (BIGDATA), 978-15386-2715-0, vol. 17, 4467-4473, 2018 (Journal Online Sources style)
11. "*Data Masking: What You Need to Know*", White Paper ,2016.
12. "*Data Masking Best Practice*", An Oracle White Paper June 2013
13. Rishabh Agarwal, Shivi Shandilya, Amandeep, Dr. Amit Singhal, "*Survey on Cloud Computing and Data Masking Techniques*", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 04, Apr-2018 .
14. Shi-Cho Cha, Kuo-Hui Yeh, "*A Data-Driven Security Risk Assessment Scheme for Personal Data Protection*", IEEE Access, VOLUME 4, 2016.
15. Ajayi, Olusola Olajide, Adebiyi, Temidayo Olarewaju, "*Application of Data Masking in Achieving Information Privacy*", IOSR Journal of Engineering (IOSRJEN), Vol. 04, Issue 02, 2278-8719, PP 13-21,2014

AUTHORS PROFILE

Akshata Khasge, is a student at Department of Computer Science and Engineering, Rashtreeya Vidyalaya college of Engineering, Bengaluru, Karnataka, India, akshatakhasge.scs19@rvce.edu.in

Dr. Nagaraja G S, is working as professor and Associate Dean at Department of Computer Science and Engineering, Rashtreeya Vidyalaya College of Engineering, Bengaluru, Karnataka, India. Published 41 international journals, 33 international conferences, 01 national journal, 22 national conferences, and 5 book chapters. ISTE presented a National award " Rajarambapu Patil" for promising engineering teacher for creative work done in Technical Education (Engineering College) for the academic year-2018.