Software As Service Attack Detection and Prevention using Manipulated QR Code

Manushree Sahay
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune-43.
sahaymanushree@gmail.com

Prof. Sandeep Vanjale
Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune-43.
sbvanjale@bvucoep.edu.in

Abstract: A sort of software service delivery paradigm known as "software as a service" (SaaS) includes a wide variety of commercial possibilities and problems. Despite being drawn to SaaS by its advantages, users and service providers are reluctant to incorporate their businesses into it because of security concerns. This article emphasizes the usefulness and adaptability of SaaS in a variety of situations, such as software defined networking, cloud computing, mobile cloud computing, and the Internet of Things. The examination of SaaS security issues, including data security, application security, and SaaS deployment security, is then started. Potential solutions or strategies that may be used in conjunction with one another are then offered for a secure SaaS platform. The SQL injection attack is the SaaS application's most dangerous vulnerability. This might result in sensitive and important data loss. Through these kinds of assaults, attackers might steal sensitive information that is crucial to a business or organization, which has a detrimental effect on both physical (like data) and intangible (like reputation) assets. This research aims to investigate the viability of using machine learning techniques for application-level SQL injection detection. Various dangerous and benign payloads were utilized to train the classifiers employed in the testing methodologies. They detect if a payload includes malicious code when given one. This study aims to identify harmful activities in a Software as a Service (SaaS) environment based on the cloud. The anti-phishing advice for this technique, which is known as a secure QR code, includes a thorough analysis of the most current research on the usability and security of QR codes. The most important use cases and accompanying attack paths were identified. To do this, we conducted a comprehensive literature study. Social engineering, or phishing, is the fraud that exploits QR codes as an attack vector most often covered in the media. The usage of QR codes on smartphones has spread from auto production plants.


1 Introduction
The cloud storage Provides a solution to the risks and challenges of cloud computing by storing data via various cloud service providers (CSPs), including vendor lock-in, and data privacy. CSB is a Software-as-a-Service (SaaS) third-party cloud storage service provider
that manages the relationship between one or more CSPs and cloud clients. Cloud is an emerging technology and cloud-based storage is a new concept that allows users not only to upload data to the internet, but also to easily access available resources and share data with anyone at any time. But cloud is a technique that generates a restore feature that enables clients to go back to a previous one attack state or Computer catastrophe provides an easy way to remove malware and computer security. The attackers have short-term windows where they should be trained and targeted for remote start-up and stoppage of VM. This is an extremely effective tool for defense. Because the hypervisor operates from Virtual Machine it is possible to control malware. VM Infrastructure will secure itself as a physical server infrastructure for such purposes.

**Collision attack:** A collision attack simply refers to the process by which an attacker utilizes one of these clashes to undermine the security that the hash was supposed to provide. Collision attack is an attack where the malicious user (i.e untrusted user) gives the wrong or misleading password about the cloud service provided by the cloud provider. In this malicious user means the user or an attacker who is never registered to the cloud service or the one who is not used the cloud service provided by the cloud provider. In this project we prevent the collision attack by avoiding the misleading password from unregistered users. Thus we check whether the user who is giving the password is registered or not. If the person giving the password is registered then he/she is called as the cloud consumer and their comments are accepted and displayed to the new user.

**Password login:** A collision is a condition whereby two messages, let say $D_1$ and $D_2$, after applying the hash value, then $H(D_1) = H(D_2)$. A collision can always be found using Brute Force algorithm, however it is computationally difficult. There are two types of collisions, the strong collision and weak collision.

**SQL injection:** SQL injection is another form of attack that occurs when SQL queries are made with user input text inserted into the query string. QR code readers are subject to data injection into their structured objects when they attempt to interpret the data of a QR code. A malicious party can create a QR code that injects arbitrary strings into a user’s data structures potentially causing harm to the user.

**SQL Injection Attack:** There are different types of SQLi attacks such as error-based, Boolean-based, time-based, and out-of-band SQLi. When exploiting an error-based SQLi vulnerability, attackers can retrieve information such as table names and content from visible database errors using the following queries:

```sql
'and(select+1+from(select+count(*),floor(rand(0)*2)from+information schema.tables+group+by+2)a)–
id=1+and(select 1 FROM(select count(*),concat((select (select concat(database())) FROM information schema.tables LIMIT 0,1),
floor(rand(0)*2))x FROM information schema.tables GROUP BY x)a)
```
Attackers can test for web application vulnerabilities to SQLi by inserting a condition into an SQL query. If the page loads as usual it indicates that the page is prone to attack. The following query is an example:

\[ \text{id}=1+\text{AND}+1=1 \]

**QR code:** There are varying sizes of QR codes with various sections designated for particular applications. Version 2 of QR codes is used in the sections that follow (Figure 1)

![QR Code](https://seer-ufu-br.online)

**Fig. 1.1 Structure of QR code Version 2**

Finder Pattern (1): The finder pattern is made up of three identical structures that may be found in all save the bottom right corner of the QR code. A 33 matrix of black modules surrounding by white modules, which are then encircled by black modules, forms the basis of each pattern. The decoder software can identify the QR code and determine its proper orientation thanks to the Finder Patterns.

Separators (2): The white separators, which are one pixel wide and help distinguish the Finder Patterns from the real data, increase their legibility.

Timing Pattern (3): The decoder software can identify the width of a single module thanks to the Timing Pattern's alternation of black and white modules.

Alignment Patterns (4): Alignment Patterns support the decoder software in compensating for moderate image distortions. Version 1 QR codes do not have Alignment Patterns. With growing size of the code, more Alignment Patterns are added.

Format Information (5): The Formation Information section consists of 15 bits next to the separators and stores information about the error correction level of the QR code and the chosen masking pattern.

Data (6): Data is converted into a bit stream and then stored in 8 bit parts (called code words) in the data section.

Error Correction (7): Similar to data codes, error correction codes are stored in 8 bit long code words in the error correction section.

Remainder Bits (8): This section consists of empty bits of data and error correction bits cannot be divided into 8 bit code words without remainder.
2 Literature Survey

According to [1] an approach which detects a query token with reserved words-based lexicon to detect SQLIA. The approach consists of two highlights: the first one creates lexicon and the second step tokenizes the input query statement and each string token was detected to predefined words lexicon to prevent SQLIA. In the detection and prevention technologies of SQL injection attacks are experimented and the result are satisfactory. This query statement is always true because it have been added by the tautology statement ( 'a'=a' ). Double dash “--” instructs the SQL parser that the rest of statement is a comment and should not be executed.

According to [2] system is online, and there is no need for implementation. This can be accessed from any location via the internet. This framework uses SQL injection protection to keep the database and data safe. It will make particular emphasis on encrypting Credit card data using AES (Advanced Encryption Standard) technique. The shop secures payment so that everything will be secured. The user pay card data is then stored in a database. The device also stores user information in an encrypted form through the use of AES encryption. The framework is designed to avoid SQL injection activities that can break into the database and run it.

According to [3] a new model for QR Code authentication and phishing detection has been presented. The proposed model will be able to detect the phishing and malicious URLs in the process of the QR Code validation as well as to prevent the user from validating it. The development of this application will help to prevent users from being tricked by the harmful QR Codes. This project is mainly for the users who scans the QR codes around them in their daily activities, such as students or employees and this tool will make sure their information will not be leaked or breached when scanning a QR code. The advantage of this detection model is it detects the abnormal and malicious links that are embedding in the QR codes to be used as a vector attack. This QR code phishing detection tool was developed and tested before the tool was presented. In order to make sure the scanning is done perfectly and the malicious and phishing page is detected, the protocols that has been used is HTTP, HTTPs, FTP, SFTP, RTSP, Telnet, NNTP, Gopher.

According to [4] an overview of the SQL injection attack and a classification of the newly proposed detection and prevention solutions. We classify the different attack sources, goals, and types. Moreover, we discuss and classify the most important and recent proposed solutions to mitigate this attack especially those based on ontology and machine learning. An overview of the SQLI attack was presented. The different attack sources, goals and types are described and discussed. There was a table that compared the various SQLI attack defenses that were suggested. We explain and debate recently proposed solutions, such as those based on ontology and machine learning.

According to [5] a SQL injection detection method that does not rely on background rule base by using a natural language processing model and deep learning framework on the basis of comprehensive domestic and international research. The method can improve the accuracy and reduce the false alarm rate while allowing the machine to automatically learn the language model features of SQL injection attacks, greatly reducing human intervention and providing some defense against 0day attacks that never occur.
According to [6] approach to visual secret sharing scheme to encode a secret QR code into distinct shares. In assessment with other techniques, the shares in proposed scheme are valid QR codes that may be decoded with some unique that means of a trendy QR code reader, so that escaping increases suspicious attackers. An existing sharing technique is subjected to loss of security. On this premise, consider the strategy for (k, n) get to structures by using the (k, k) sharing occurrence on each k-member subset dependent on specific relationship. In addition, the secret message is recovered with the aid of XOR-ing the qualified shares. This operation which can effortlessly be achieved the use of smartphones or different QR scanning gadgets. Contribution work is, working on optimal partitioning methods and compare original message with shared message using hashing techniques.

According to [7] Quick response code phishing attack is when an attacker lures its victims to voluntarily divulge personal information such as password, personal identification number, username and other information such as online banking details through the use of quick response code. This attack is on the rise as more and more people have adopted mobile phone usage not just for communication only but to perform transaction seamlessly. The ease of creation and use of quick response code has made it easily acceptable to both provider of goods and services and consumers. This attack is semantic as it exploits human vulnerabilities; as users can hardly know what is hidden in the quick response code before usage. This study reviewed various methodologies that earlier researcher have used to detect this semantic-based attack of phishing. The strength of each methodology, its weakness and general research gaps identified.

According to [8] main objective is to detect malicious URLs embedded in QR codes. A dataset of 90,000 benign and malicious URLs was collected from various resources, and their lexical properties were extracted. Two computational intelligence models, fuzzy logic and multilayer perceptron artificial neural network (MLP-ANN), were applied and compared. An MLP-ANN was identified as the best classifier for detecting malicious URLs, and a proactive, secure, real-time computational intelligence barcode scanner implementation (BarCI) against malicious QR code links was proposed based on this classifier.

According to [9] a detailed review on various types of SQL injection attacks and detection techniques based on machine learning. We also propose future expectations and possible development of countermeasures against SQL injection vulnerability at the end of the article. Structured Query Language (SQL) provide a way to manipulate data and change database structure. If website use it unsafely, then maybe there are SQL injection vulnerabilities in this website. SQL injection attacks allow hacker get access to sensitive information without authentication and authorization, therefor it can make serious harm to the website with low cost and threshold.

According to [10] the work focuses on extracting SQL injection patterns with the help of existing parsing and tagging techniques. The uses pattern-based Neural Network model for SQL injection detection working on a simple and efficient method. Objective of this work is to find the pattern of the WHERE clause of the SQL query using natural language tagging techniques to generalize the patterns of legitimate and injected queries.

According to [11], Big data and Internet of Things (IoT)-based applications are employed in smart environments. These industries aim to identify important application areas, current
trends, data formats, and on-going difficulties. To our knowledge, it is the first systematic research of its kind, looking at academic articles published in peer-reviewed venues between 2011 and 2019, utilizing a four-step selection procedure of identification, screening, eligibility, and inclusion. They conducted a systematic study and addressed six major research issues to explore these data. The results suggest that merging big data and IoT technologies opens up new opportunities for smart environment applications that monitor, conserve, and improve natural resources in the real world. This research looks at smart environment monitoring, smart farming/agriculture, smart metering, & smart disaster alerts.

In improved Honeypot cryptographic technique for cloud security prediction, A. Mondal and R. T. Goswami [12] suggested that data must be protected against infiltration or other kinds of attack. Their technique of privacy protection uses a cryptographic mechanism. Encryption is done using the Honeypot algorithm. When a data owner requests a file, the cloud server generates a key and verifies it with the user. When the user gives the key, the file is decrypted and delivered to the user.

According to[13] a cooperative method to identify and stop assaults on Software-Defined Networks (SDN) using Distributed Denial-of-Service (DDoS) floods. For the purpose of detecting DDoS traffic flows in an SDN controller, this method merges the sflow-RT application with Snort rules. Multiple Ryus SDN controllers can share DDoS detection and mitigation rules using Redis Simple Message Queue (RSMQ). The rule-sharing enables the controller's processing overhead for DDoS detection and mitigation to be reduced. According to the experimental findings, DDoS assaults may be considerably detected and prevented across multi-controller domains by employing the RSMQ technique. Additionally, it offers DDoS early detection and mitigation at reduced controller overhead.

According to [14] It is crucial to identify assaults that result in Cloud services being unavailable since DDOS attack detection has increased in frequency in dispersed environments like the cloud. Machine learning models may be used to train and evaluate attack detection datasets in order to recognize such assaults. As an alternative, we may use the regression analysis method by using multiple linear regression analysis, one of its crucial forms. The goal of this study's research is to create a machine learning model that combines feature selection with regression analysis and information acquisition. The goal of the research is to investigate the issue of DDoS attack detection in a cloud environment by taking into account the most well-liked CICIDS 2017 benchmark dataset and using multiple regression analysis to create a machine learning model to anticipate DDoS and Bot attacks by taking into account a Friday afternoon traffic logfile.

According to [15] One of the main dangers to online applications is SQLIA. In order to maintain confidentiality and integrity, web applications must safeguard their database from a variety of risks. In SQLIA, hackers are able to attack the system using a specially crafted query statement through a web input form, steal identities, gain access to private data, and manipulate existing data, all of which can have a variety of disastrous consequences. This essay discusses SQLIA prevention strategies and tactics. The suggested method is used to identify SQL injection, block it, and provide results that are appropriate.
3 Proposed System Details

3.1 Problem Statement
The system for searching keywords using similarity-based techniques on encrypted data also classification the document based on weight and query. The system is designed for SaaS (Software-as-a-service) Attack detection and prevention of malicious activities in the cloud environment.

3.2 Objective
- To study and analysis software as services attack detection and prevention using QR code generation and machine learning techniques.
- To design and develop a method for dynamic QR code generation for heterogeneous network attacks on web applications.
- To design and develop a system for various Attack detection with internal as well as external attacks like collision attack, SQL injection attack, etc.
- To explore and validate the proposed system result with various existing systems and the effectiveness of proposed model.

3.3 System Architecture
We propose a protected information sharing plan for element individuals. Initially, we propose a protected path for key dissemination with secure correspondence channels, and the clients can safely acquire their public keys from the gathering chief. Our proposed system uses three different entities: data owner, user, cloud server, and the attacker are an untrusted entity. In this module first data owner upload the data file to the cloud server using a cryptography algorithm. Once data has been stored in the database, the owner successfully gets the notification about file storage. The data owner has full access to a specific data file he can share or access so that the data owner can share any file with any group manager. Then it will automatically access all group members. The shared group members can access each file anytime by the cloud server. If he can try to generate any collision attack using SQL injection queries, even our system will prevent such attacks. The overall approach improves the system efficiency as well security on a drastic level.

Figure 2: Proposed System architecture
Login and Registration Module: In this module user can make registration, once registration has done by user, then he/she will get access for login in system. Any authenticate user can communicate with system anywhere, anytime with the help of cloud.

Data Uploading: In first phase once data owner uploads the file. In that module data encryption done using PBEWithMD5AndDES and SHA256 encryption scheme and the same time keys send to EC2 Cloud. Data Owner will upload file for his backup and can allow the file to be accessed by friend user. He will be able to access the uploaded file by entering proper credentials sent to his registered email id. Files uploaded by user will get scan by algorithm and if similar contents are there in old and current new file then previous file will get stored. Common files in which all contents are exactly similar if gets uploaded on the system by multiple users will get registered as the first owners file and others can access it as friend user which will avoid duplication of file at server.

Encryption and Decryption: Decryption is the technique used to change an unreadable message into a legible and intelligible one. Encryption is the method used to change an undamaged communication (plaintext) into an unreadable one (chipper text). One or more cryptographic keys control the encryption and decryption process. A cryptosystem is a tool that allows plaintext to be converted to chipper text and vice versa. Cryptography may be separated into symmetric key cryptography (also known as symmetric-key cryptography) and asymmetric key cryptography (also known as asymmetric-key cryptography) depending on the keys used for encryption and decryption.

Data Sharing: In that phase data sharing done by data owner, he can any file to any user in cloud group. Friend user can access the shared file to him by data owner by following the login process and proper credentials send to him for that particular file.

Access Control and revocation: In access control any user can view or access the file shared by user to him. In revocation data owner can revoke the file access to specific user. The common files uploaded by multiple users can’t be deleted will get deleted according to maximum requirement time specified by out of multiple users.

File request and download: user can give the download request to cloud server, at the same Data Owner at verification has done.

Cloud Storage Service Provider (CSP)
Cloud Storage Services Provider provides database. It allows data owner to keep any kind of information. CSP also allows to the user to make the user defined database schema. According to user requirement the space for the user instance will be allocated by CSP
3.4 Algorithm Design

**Similarity (Machine learning) Algorithm:**

**Input:** Normative training set Testset using normalized data from Train_Data The threshold Th_values was specified by Test_Data.

**Output:** Output set using the parameters {Predicted_class, weight_Score}

**Step 1:** To verify training rules, read all test data from Test_Data[] using the function below. The data is then normalized and altered to meet the needs of the algorithms.

\[
\text{test\_Feature(data)} = \sum_{m=1}^{n} (\text{Attribute\_Set}[A[m] ... A[n]} \leftarrow \text{Test\_Data})
\]

**Step 2:** choose the features from the test's extracted attributes set. With the code below, create a feature map using the data as features.

\[
\text{Test\_FeatureMap [t.....n]} = \sum_{x=1}^{n} (t) \leftarrow \text{test\_Feature(x)}
\]

Test\_FeatureMap [x] are the selected features, and those selected features are stored in Test\_FeatureMap

**Step 3:** Now read the complete targeting dataset to create the hidden layer for the sense layer's categorization of all test data.

\[
\text{train\_Feature(data)} = \sum_{m=1}^{n} (\text{Attribute\_Set}[A[m] ... A[n]} \leftarrow \text{Train\_Data})
\]

**Step 4:** Create the training map from the input dataset using the function below.

\[
\text{Train\_FeatureMap [t.....n]} = \sum_{x=1}^{n} (t) \leftarrow \text{train\_Feature(x)}
\]

Train\_FeatureMap[t] is the hidden layer map that generates feature vector for build the hidden layer. That evaluates the entire test instances with train data.

**Step 5:** After generating the feature map we calculate similarity weight.

\[
\text{Gen\_weight} = \text{CalcWeight} (\text{Test\_FeatureMap} || \sum_{i=1}^{n} \text{Train\_FeatureMap[i]})
\]

**Step 6:** Evaluate the current weight with desired threshold

\[
\text{if(Gen\_weight > qTh)}
\]

**Step 7:** Out\_List.add (trainF.class, weight)

**Step 8:** Go to step 1 and continue when Test\_Data == null

**Step 9:** Return Out\_List.

**Algorithm 2: PBEWithMD5AndDES (Encryption and Decryption) Algorithm**

The Message Digest 5 (MD5) and Data Encryption Standard (DES) algorithms are used in the cryptographic technique known as PBE with MD5 and DES. MD5 creates a 128 bit message digest from messages of any length.

**Key Generation Process**

**Step 1:** Char [] = char.random [5];
**Step 2:** string Key = (string) char []
**Step 3:** Return Key

**Encryption Process**

**Input:** plain text p, and private key k

**Output:** cipher text C
Step 1: Generate instance of PBEWithMD5AndDES
Step 2: Set encrypt mode with cipher instance.
Step 3: Change byte [] plaintext = Plain byte []
Step 4: [] enc = apply cipher method on (plaintext, k)
Step 5: Encstring = apply 64 base encoder on [] enc
Step 6: return Encstring

**Decryption Process**

**Input:** cipher text C, key k

**Output:** Plain text p

Step 1: Set k as private key for decryption.
Step 2: Set decrypt mode with cipher instance.
Step 3: byte [] ks = 64 base decoder on (c)
Step 4: byte [] utf = apply decipher method on (ks, k)
Step 5: plain = convert into string class (utf)
Step 6: return plain

**Algorithms 3: Role Based Access Control Algorithms:**

**Input:** Attribute Email-ID, File Data and File key_data.

**Output:** Rule set as policies or signatures.

Step 1: Prepare the data string S_list []
Step 2: Prepare a = 0, k = 0, User Email-ID
Step 3: Read Filedata and filekey

a ← {filekey list [i . . . . . . n]}
k ← {Email-ID List [i . . . . . . n]}

Step 3: for each (read a to S_list)
If (key_data. Equals (a) && User Email-ID. Equals (k))
Then User File Share information show
Else
Then User File Not Share information show
End for
Step 4: End Procedure

**4 Mathematical Models**

First we consider a

Act= {Act1, Act2, At3, . . . , An} each collection contains the system's unique module operation.

Act1= {file data uploading phase or file Download phase}
Act2= {data encryption phase}
Act3= {User activity}
Act4= {User block and unblock}

Act1 identify the first module, which is the user's ability to upload documents.

\[ \text{Data}[doc] = \text{doc}[k] + \sum_{i=k=0}^{n} (a01, a02 ....... an) \]

Doc[k] ← {Attri1, Attrii2……Attrin.} each documents contains the set of attributes
The formula below has been used to describe role-based access control for each ith user

\[ U[i]_{\text{file}(x)} = \sum_{n=1}^{m} (u[n]_{\text{read, write}}) \]

5 Results And Discussion

Performance comparison of proposed and existing methods

<table>
<thead>
<tr>
<th>Attack Type</th>
<th>No. of input values</th>
<th>Correct Detection</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Injection</td>
<td>140</td>
<td>138</td>
<td>98.57</td>
</tr>
<tr>
<td>Collusion</td>
<td>150</td>
<td>149</td>
<td>99.33</td>
</tr>
</tbody>
</table>

The Table 1 describes detection accuracy for 2 different attacks such as SQL injection and Collision attacks. The number of malicious inputs and correctly detected by algorithm are mentioned using attribute 2 and 3.

Figure 3: attack detection with various input types

The Figure 2 describes performance evaluation of proposed algorithms vs conventional attack detection techniques.

Figure 4: comparative analysis of proposed attack detection methods vs traditional methods
The SQL injection achieves similar results when evaluate with existing methods but collision attacks improvise 14% over the exiting methods.

6 Conclusions
In this work, we have proposed a novel approach to enable digital forensics in the cloud environment with respect to performance by taking virtual machine details (i.e., IP address and mac address) as evidence. The approach incorporates intrusion detection system in virtual machine to identify the malicious virtual machine and improves the cloud performance in terms of size and time by storing details of malicious virtual machine. The proposed approach takes details of suspected virtual machine and stored in persistent storage, hence improves the performance of cloud. A quick and efficient approach of detecting SQL injection threats. When arguments are provided, the technique extracts the value from a SQL query attribute of web pages and compares it to a specified value. This approach combines static analysis with dynamic analysis. In this paper we outlined the dangers of possible attacks utilizing manipulated QR codes. Since QR codes gain increasing popularity through their use for marketing purposes, we expect that this kind of attack will receive more and more attention by the hacking community in the future.

7 Future Works
To evaluate the proposed system on various distributed environment in fog nodes with different input objects.

References