SQL Injection Detection and Prevention Techniques

Atefeh Tajpour, Suhaimi Ibrahim, Maslin Masrom
University Technology Malaysia, tajpour81sn@yahoo.com, suhaimiibrahim@utm.my, maslin@ic.utm.my
doi: 10.4156/ijact.vol3.issue7.11

Abstract

SQL injection is a type of attack which the attacker adds Structured Query Language code to a web form input box to gain access or make changes to data. SQL injection vulnerability allows an attacker to flow commands directly to a web application's underlying database and destroy functionality or confidentiality. Researchers have proposed different tools to detect and prevent this vulnerability. In this paper we present SQL injection attack types and also current techniques which can detect or prevent these attacks. Finally we evaluate these techniques.

Keywords: SQL Injection Attacks, Detection, Prevention, Evaluation, Technique, Web Application Security.

1. INTRODUCTION

Web applications are often vulnerable for attackers easily access to the application's underlying database. SQL injection attack occurs when a malicious user, through specifically crafted input, causes a web application to generate and send a query that functions differently than the programmer intended.

SQL Injection Attacks (SQLIAs) have known as one of the most common threats to the security of database-driven applications. So there is not enough assurance for confidentiality and integrity of this information. SQLIA is a class of code injection attacks that take advantage of lack of user input validation. In fact, attackers can shape their illegitimate input as parts of final query string which operate by databases. Financial web applications or secret information systems could be the victims of this vulnerability because attackers by abusing this vulnerability can threat their authority, integrity and confidentiality. So, developers addressed some defensive coding practices to eliminate this vulnerability but they are not sufficient.

For preventing the SQLIAs, defensive coding has been offered as a solution but it is very difficult. Not only developers try to put some controls in their source code but also attackers continue to bring some new ways to bypass these controls. Hence it is difficult to keep developers up to date, according the last and the best defensive coding practices. On the other hand, implementing of defensive coding is very difficult and need to special skills and also erring. These problems motivate the need for a solution to the SQL injection problem.

Researchers have proposed some tools to help developers to compensate the shortcoming of the defensive coding [7, 10, 12]. The problem is that some current tools could not address all attack types or some of them need special deployment requirements.

2. OVERVIEW OF SQL INJECTION ATTACK

2.1. Definition of SQLIA

Most web applications today use a multi-tier design, usually with three tiers: a presentation, a processing and a data tier. The presentation tier is the HTTP web interface, the application tier implements the software functionality, and the data tier keeps data structured and answers to requests from the application tier [31]. Meanwhile, large companies developing SQL-based database management systems rely heavily on hardware to ensure the desired performance [32].

SQL injection is a type of attack which the attacker adds Structured Query Language code to input box of a web form to gain access or make changes to data. SQL injection vulnerability allows an
attacker to flow commands directly to a web application's underlying database and destroy functionality or confidentiality.

2.2. SQL Injection Attacks (SQLIAs) Process

SQLIA is a hacking technique which the attacker adds SQL statements through a web application's input fields or hidden parameters to access to resources. Lack of input validation in web applications causes hacker to be successful. For the following examples we will assume that a web application receives a HTTP request from a client as input and generates a SQL statement as output for the back end database server.

For example an administrator will be authenticated after typing: employee id=112 and password=admin. Figure 1 describes a login by a malicious user exploiting SQL Injection vulnerability [11]. Basically it is structured in three phases:

1. an attacker sends the malicious HTTP request to the web application
2. creates the SQL statement
3. submits the SQL statement to the back end database

![Figure 1. Example of a SQL Injection Attack](image)

The above SQL statement is always true because of the Boolean tautology we appended (OR 1=1) so, we will access to the web application as an administrator without knowing the right password.

2.3. CONSEQUENCE OF SQLIA

The results of SQLIA can be disastrous because a successful SQL injection can read sensitive data from the database, modify database data (Insert/Update/Delete), execute administrative operations on the DB (such as shutdown the DBMS), recover the contents on the DBMS file system and execute commands (xp_cmdshell) to the operating system. The main consequences of these vulnerabilities are attacks on [11]:

**Authorization:** Critical data that are stored in a vulnerable SQL database may be altered by a successful SQLIA, as authorization privilege.

**Authentication:** If there is no any proper control on username and password inside the authentication page, it may be possible to login to a system as a normal user without knowing the right username and/or password.

**Confidentiality:** Usually databases are consisting of sensitive data such as personal information, credit card numbers and/or social numbers. Therefore loss of confidentiality is a big problem with SQL Injection vulnerability. Actually, theft of sensitive data is one of the most common intentions of attackers.

**Integrity:** By a successful SQLIA not only an attacker reads sensitive information, but also, it is possible to change or delete this private information.
3. CLASSIFICATION OF SQLIA

Depending on many factors, SQL Injection attacks can be divided into several groups. Here, there are two categories in which are possible to unite all the SQLIAs. This classification is according to the main publications related to the phenomena of SQL Injection [2].

3.1. By Intent

An important classification of SQLIA is related to the attacker's intent, or in other words, the goal of the attack [4].

- **Extracting data** - This category of attacks tries to extract data values from the back end database. Based on the type of web application, this information could be sensitive, for example, credit card numbers, social numbers, private data are highly valuable to the attacker. This kind of intent is the most common type of SQLIA.

- **Adding or modifying data** - The purpose of these attacks is to add or change data values within a database.

- **Performing database finger printing** – In this category of attack the malicious user wants to discover technical information on the database such as the type and version that a specific web application is using. It is noticeable that certain types of databases respond differently to different queries and attacks, and this information can be used to "fingerprint" the database. Once the intruder knows the type and the version of the database it is possible to organize a particular attack to that database.

- **Bypassing authentication** - By this attack, intruders try to bypass database and application authentication mechanisms. Once it has been over passed, such mechanisms could allow the intruder to assume the rights and privileges associated with another application user.

- **Identifying inject able parameters** - Its goal is to explore a web application to discover which parameters and user-input fields are vulnerable to SQLIA. By using an automated tool called a "vulnerabilities scanner" this intent can be identified.

- **Determining database schema** - The goal of this attack is to obtain all the database schema information (such as table names, column names, and column data types). This is very useful to an attacker to gather this information to extract data from the database successfully. Usually by exploiting specific tools such as penetration testers and vulnerabilities scanners this goal is achieved.

- **Evading detection** – in this category attackers try to avoid auditing and detection, including evading defensive coding practices and also many automated prevention techniques. These attack techniques try to bypass Intrusion Detection and Prevention or other security mechanisms which provide security for systems.

- **Performing denial of service** – in this category intruders make interrupt in system services by performing some instruction so the database of a web application shutdown, thus denying service happen. Attacks involving locking or dropping database tables also fall into this category.

- **Executing remote commands** - The goal of this action is to execute arbitrary commands on the database. These commands can be stored procedures or functions available to database users. This kind of attack is the most dangerous because it may allow the intruder to gain control of the whole system. For instance Microsoft's SQL Server supports a stored procedure xp - cmdshell that permits to arbitrary command execution, and if this attack runs then complete compromise of the server is unavoidable.

3.2. By Type

There are different methods of attacks. Depending on the goal of attacker are performed together or sequentially. For a successful SQLIA the attacker should append a syntactically correct command to the original SQL query. Now the following classification of SQLIAs in accordance to the Halfond, Viegas, and Orso researches [4, 11] will be presented.
Tautologies: This type of attack injects SQL tokens to the conditional query statement to be evaluated always true.

Illegal/Logically Incorrect Queries: when a query is rejected, an error message is returned from the database including useful debugging information. This error messages help attacker to find vulnerable parameters in the application and consequently database of the application.

Union Query: By this technique, attackers join injected query to the safe query by the word UNION and then can get data about other tables from the application.

Piggy-backed Queries: In this type of attack, intruders exploit database by the query delimiter, such as ";", to append extra query to the original query. With a successful attack database receives and execute a multiple distinct queries. Normally the first query is legitimate query, whereas following queries could be illegitimate.

Stored Procedure: Stored procedure is a part of database that programmer could set an extra abstraction layer on the database. As stored procedure could be coded by programmer, so, this part is as inject able as web application forms. Depend on specific stored procedure on the database there are different ways to attack.

Alternate Encodings: In this technique, attackers modify the injection query by using alternate encoding, such as hexadecimal, ASCII, and Unicode. Because by this way they can escape from developer’s filter which scan input queries for special known "bad character". For example attacker use char (44) instead of single quote that is a bad character.

Inference: By this type of attack, intruders change the behaviour of a database or application. There are two well-known attack techniques that are based on inference: blind-injection and timing attacks.

- Blind Injection: Sometimes developers hide the error details which help attackers to compromise the database. So the SQLIA would be more difficult but not impossible. An attacker can still steal data by asking a series of True/False questions through SQL statements.

- Timing Attacks: A timing attack lets an attacker gather information from a database by observing timing delays in the database's responses. This technique uses an if-then statement for injecting queries. WAITFOR is a keyword along the branches, which causes the database to delay its response by a specified time.

4. PREVENTION OF SQL INJECTION ATTACK.

Researchers have proposed a wide range of techniques to address the problem of SQL injection. These techniques range from development best practices to fully automated frameworks for detecting and preventing SQLIAs [1].

Huang and colleagues [18] propose WAVES, a black-box technique for testing web applications for SQL injection vulnerabilities. The tool identify all points a web application that can be used to inject SQLIAs. It builds attacks that target these points and monitors the application how response to the attacks by utilize machine learning.

JDBC-Checker [12, 13] was not developed with the intent of detecting and preventing general SQLIAs, but can be used to prevent attacks that take advantage of type mismatches in a dynamically-generated query string. As most of the SQLIAs consist of syntactically and type correct queries so this technique would not catch more general forms of these attacks.

Wassermann and Su propose Taoutology Cheker [23] that uses static analysis to stop tautology attack. The important limitation of this technique is that its scope is limited to tautology and cannot detect or prevent other types of attacks.

Xiang Fu and Kai Qian [28] proposed the design of a static analysis framework, called SAFELI for identifying SQLA vulnerabilities at compile time. SAFELI statically monitor the MSIL (Microsoft Symbolic intermediate language) byte code of an ASP.NET Web application, using symbolic execution. SAFELI can analyze the source code and will be able to identify delicate vulnerabilities that cannot be discovered by black-box vulnerability scanners. The main drawback of this technique is that this approach can discover the SQL injection attacks only on Microsoft based product.

CANDID [2, 7] modifies web applications written in Java through a program transformation. This tool dynamically mines the programmer-intended query structure on any input and detects attacks by
comparing it against the structure of the actual query issued. CANDID’s natural and simple approach turns out to be very powerful for detection of SQL injection attacks.

In SQL Guard [10] and SQL Check [5] queries are checked at runtime based on a model which is expressed as a grammar that only accepts legal queries. SQL Guard examines the structure of the query before and after the addition of user-input based on the model. In SQL Check, the model is specified independently by the developer. Both approaches use a secret key to delimit user input during parsing by the runtime checker, so security of the approach is dependent on attackers not being able to discover the key. In two approaches developer should to modify code to use a special intermediate library or manually insert special markers into the code where user input is added to a dynamically generated query.

AMNESIA combines static analysis and runtime monitoring [16, 17]. In static phase, it builds models of the different types of queries which an application can legally generate at each point of access to the database. Queries are intercepted before they are sent to the database and are checked against the statically built models, in dynamic phase. Queries that violate the model are prevented from accessing to the database. The primary limitation of this tool is that its success is dependent on the accuracy of its static analysis for building query models.

WebSSARI [15] use static analysis to check taint flows against preconditions for sensitive functions. It works based on sanitized input that has passed through a predefined set of filters. The limitation of approach is adequate preconditions for sensitive functions cannot be accurately expressed so some filters may be omitted.

Livshits and Lam [27] use static analysis techniques to detect vulnerabilities in software. Java Static Tainting uses information flow techniques to detect when tainted input has been used to make a SQLIA. The primary limitation of this approach is that it can detect only known patterns of SQLIAs and it can generate a relatively high amount of false positives because it uses a conservative analysis.

Java Dynamic Tainting [21] and SecuriFly [14] is another tool that was implemented for java. Despite of other tool, chase string instead of character for taint information and try to sanitize query strings that have been generated using tainted input but unfortunately injection in numeric fields cannot stop by this approach. Difficulty of identifying all sources of user input is the main limitation of this approach.

Two similar approaches by Nguyen-Tuong [20] and Pietraszek [19], modify a PHP interpreter to track precise per-character taint information. A context sensitive analysis is used to detect and reject queries if certain types of SQL tokens has been constructed by illegitimate input. Limitation of these two approaches is that they require rewriting code.

Two approaches, SQL DOM [25] and Safe Query Objects [24], use database queries encapsulation for trustable access to databases. They use a type-checked API which cause query building process be systematic. Consequently by API they apply coding best practices such as input filtering and strict user input type checking. The drawback of the approaches is that developer should learn new programming paradigm or query-development process.

Positive tainting [1] not only focuses on positive tainting rather than negative tainting but also it is automated and does need developer intervention. Moreover this approach benefits from syntax-aware evaluation, which gives developers a mechanism to regulate the usage of string data based not only on its source, but also on its syntactical role in a query string.

IDS [6] use an Intrusion Detection System (IDS) to detect SQLIAs, based on a machine learning technique. The technique builds models of the typical queries and then at runtime, queries that do not match the model would be identified as attack. This tool detects attacks successfully but it depends on training seriously. Else, many false positives and false negatives would be generated.

Another approach in this category is SQL-IDS [8] which focus on writing specifications for the web application that describe the intended structure of SQL statements that are produced by the application, and in automatically monitoring the execution of these SQL statements for violations with respect to these specifications.

A proxy filtering system that intensifies input validation rules on the data flowing to a Web application is called Security Gateway [26]. In this technique for transferring parameters from webpage to application server, developers should use Security Policy Descriptor Language (SPDL). So developer should know which data should be filtered and also what patterns should apply to the data.
SQLPrevent [11] consists of an HTTP request interceptor. The original data flow is modified when SQLPrevent is deployed into a web server. The HTTP requests are saved into the current thread-local storage. Then, SQL interceptor intercepts the SQL statements that are made by web application and passes them to the SQLIA detector module. Consequently, HTTP request from thread-local storage is fetched and examined to determine whether it contains an SQLIA. The malicious SQL statement would be prevented to be sent to the database, if it is suspicious to SQLIA.

Swaddler [3] analyzes the internal state of a web application. It works based on both single and multiple variables and shows an impressive way against complex attacks to web applications. First, the approach describes the normal values for the application’s state variables in critical points of the application’s components. Then, during the detection phase, it monitors the application’s execution to identify abnormal states.

In SQLrand [22] instead of normal SQL keywords developers create queries using randomized instructions. In this approach a proxy filter intercepts queries to the database and de-randomizes the keywords. By using the randomized instruction set, attacker’s injected code could not have been constructed. As it uses a secret key to modify instructions, security of the approach is dependent on attacker ability to seize the key. It requires the integration of a proxy for the database in the system same as developer training.

5. EVALUATION

In this section, the SQL injection detection or prevention techniques presented in section V would be compared. It is noticeable that this comparison is based on the evaluation which the authors of techniques have done, not empirically experience, because most of techniques are not available. It is noticeable that Halfond, Viegas, and Orso [4] have done a complete evaluation for SQL injection detection techniques in 2006 but this paper covers more techniques after 2006.

5.1. Comparison of SQL Injection Detection/Prevention Techniques Based on Attack Types

Proposed techniques were compared to assess whether it was capable of addressing the different attack types presented in section III. It is noticeable that this comparison is based on the articles not empirically experience.

Tables 1 summarize the results of this comparison. The symbol “●” is used for technique that can successfully stop all attacks of that type. The symbol “−” is used for technique that is not able to stop attacks of that type. The symbol “○” refers to technique that stop the attack type only partially because of natural limitations of the underlying approach.

Table 2, illustrates the addressing percentage of SQL Injection attacks among SQL Injection detection or prevention techniques. The percentage of techniques that stop Tautology is calculated by this formula:

\[
\frac{\text{Number of techniques that can stop Tautology}}{\text{Total number of techniques}} \times 100 = \frac{14}{23} \times 100 = 61
\]
It is evident that only 26% of techniques can stop Stored Procedure. Nevertheless it is unfortunate that 39% of techniques cannot stop this type of attack and Alter Encodings is in the second grade with 48% of stopping and 17% of non stopping by detective or preventive techniques. It is interesting that almost steadily, 35% of attack types could be addressed by these techniques partially.

<table>
<thead>
<tr>
<th>Attack Types</th>
<th>Techniques that can stop all attacks of that type(s)</th>
<th>Techniques that address the attack type only partially(*)</th>
<th>Techniques that cannot stop attacks of that type(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taut</td>
<td>%61</td>
<td>%35</td>
<td>%4</td>
</tr>
<tr>
<td>Illegal/Incorrect</td>
<td>%56</td>
<td>%35</td>
<td>%9</td>
</tr>
<tr>
<td>Piggy-back</td>
<td>%61</td>
<td>%35</td>
<td>%4</td>
</tr>
<tr>
<td>Union</td>
<td>%61</td>
<td>%35</td>
<td>%4</td>
</tr>
<tr>
<td>Stored Proc</td>
<td>%26</td>
<td>%35</td>
<td>%39</td>
</tr>
<tr>
<td>Infer</td>
<td>%61</td>
<td>%35</td>
<td>%4</td>
</tr>
<tr>
<td>Alter Encodings</td>
<td>%48</td>
<td>%35</td>
<td>%17</td>
</tr>
</tbody>
</table>

5.2. Comparison of SQL Injection Detection/Prevention techniques Based on Deployment Requirement

Each technique with respect to the following criteria was evaluated: (1) Does the technique require developers to modify their code base? (2) What is the degree of automation of the detection aspect of the technique? (3) What is the degree of automation of the prevention aspect of the technique? (4) What infrastructure (not including the technique itself) is needed to successfully use the technique? The results of this classification are summarized in Table 3.

Table 3 determines the degree of automation of technique in detection or prevention of attacks and also it could be cleared that which technique needs to modify the source code of application. Moreover, additional infrastructure that is required for each technique is illustrated.
6. ACKNOWLEDGMENT

This project is sponsored by the UTM-RU Grant under the Vot. No. 00H68. Thanks to UTM-RMC and other individuals who are directly or indirectly involved in this project.

7. CONCLUSION AND FUTUREWOK

In this paper we first identified the various types of SQLIAs. Then we investigated on SQL injection detection and prevention techniques. After that we compared these techniques in terms of their ability to stop SQLIA. Regarding the results, some current techniques’ ability should be improved for stopping SQLI attacks. Moreover, we compared these approaches in deployment requirements that lead to inconvenience for users.

In our future work we separate techniques which have been implemented as tool, then compare effectiveness, efficiency, stability, flexibility and performance of tools to show the strength and weakness of the tool.

8. REFERENCES


