A Survey of SQL Injection Attack Using Machine Learning

Riya Patel, Dr. Manish Patel
M.Tech in Cyber Security
Sankalchand Patel University

Abstract *--In today's world, we all are depending on web application. Numbers of users for web application is also increased day by day. Most of organization use database to store data related to users or information which users are served. Structured Query Language is widely used to communicate with database. In SQL Injection attack malicious SQL statement is executed on the database by the attacker. SQL injection is very serious security threat as it can be done to steal the content of database; change the values stored in database even whole database can be erased. Using SQL injection attackers can gain unauthorized access to the entire database. SQL Injection attacks are possible when web application do not properly validate or filter data entered by user. There are various machine learning algorithm used to not only detect but also analyze the SQL injection threat as naïve byes, Gradient Boosting, Support Vector Machine, Decision Tree, Convolution Neural Network.*

Keywords **-- Structured Query Language injection, Detection, Web Security, Machine Learning**

1. Introduction

– Thousands of organizations around the world use website as a medium for daily transactions and business. Data for each website is stored in a relational database powered by instructions written in a special language called Structured Query Language (SQL). Most databases are vulnerable to cyber threats, most commonly facing SQL injection. SQL injection is a method of stealing data from your backend. According to the OWASP, SQL injection attacks are techniques used in hacking or cracking to gain access to information or unauthorized activity. Using SQL injection, an attacker could exploit vulnerability on her website to gain access to administrator credentials. There are common vulnerabilities. Open ports, weak passwords, weak firewalls, outdated software, invalidated user input, etc. Therefore, it's important to make your website intelligent enough to detect when input comes from bogus or malicious sources.

Attackers use SQL injection to find user credentials in the database. This user can be an administrator with all privileges. By exploiting this vulnerability, hackers can gain full access to all data in the database. SQL injection is used to change existing data and insert new data to the database. For example, in financial applications, hackers can use SQL injection attacks to modify account balances, transactions, and money transfer information. You can also use SQL injection to delete records from your database or delete tables. If administrators perform database backups, deleting data can impact application availability .A SQL injection attack is a hacking technique in which an attacker injects SQL statements through input fields or hidden parameters in a web application to access resources. Web applications do not validate input, so a hacker could successfully send malicious queries to the database server. The database executes this malicious query without your knowledge.

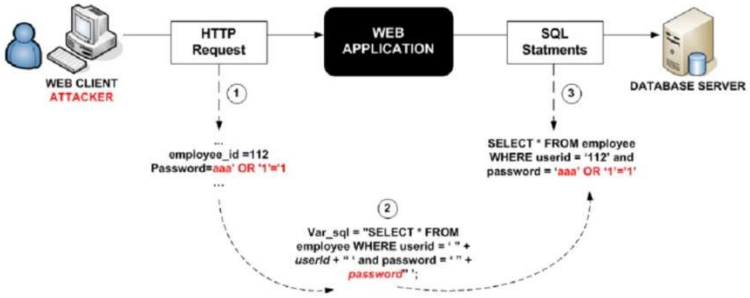
Hackers can then read confidential data from the database, modify (insert/update/delete) database data, and perform administrative operations on the database. SQLIA allow hackers to spoof, manipulate existing data, because denial problems like invalidating transaction or changing balance, allowing full disclosure of all data in the system, destroying data or you can destroy it inaccessible and become an administrator of the database server.

1.1 SQLI Attack Process:

In order to access resources, an attacker using the SQLIA hacking technique inserts SQL statements into a web application's input fields or hidden parameters. Web applications' lack of input validation makes it easier for hackers to succeed. In the examples that follow, we'll assume that a web application gets a client's HTTP request as input and outputs a SQL statement for the back-end database server. For

instance, after entering the credentials "employee id=112 and password=admin," an administrator will be verified. Image narrates a login by malicious user exploiting SQLI vulnerability. It has three phases:

a. Attacker sends the malicious HTTP Request to the web application
b. Generate the SQL statement
c. Submit the SQL query to back end database.



The above SQL query is always true because we use OR logic so, we can access web application as admin without knowing right password.

1.2 Types of SQLIA:

a. Tautologies:

To perform SQL injection, tautology-based attacks use one or more conditional SQL statement queries such as (1=1) or (- -). This method is most frequently used to gain access to databases by avoiding authentication on web page. For Example query for login:

Select *from emp where id= 'hyji' and pwd=123 or 1=1;

With the help of above query attacker get all records of table because WHERE clause always return true by using 1=1 whether pwd is correct or not.

b. Logically Incorrect queries:

The logically incorrect attack makes use of the error messages that the database returns for a wrong query. These database error reports frequently include helpful information that enables attackers to identify susceptible application parameters and database structure. For Example:

Select *from user_info where username='' group by user id having 1=1 and pwd='1234560'

This query generate error message like column user_info.usename is invalid and displays the table with column name. In this way all column name can be extracted.

c. Piggy-Backed Queries:

A form of attack known as "piggy-backed queries" attacks a database by inserting extra query statements into the original query using a query delimiter like ";". In this approach, the initial query is original, whereas the succeeding is injected. This type is serious because it allows an attacker to add any kind of SQL query. Example:

Select pass from tab where id=1 and password=0; drop table tab

d. Union Query:

In this attack, attacker inserts additional statement into the original SQL statement. This can be done by inserting either a UNION query or a statement in the form "<SQL statement>" into vulnerable parameter. This attack caused the database to return a dataset consisting of the combined results of the original query and the injected query.

Select name from user where id=1 union all select ph_no from contact;

This query returns a combined result of the original and injected query.

e. Stored Procedures:

The attacker's focus with this technique is on the database system's stored procedures. Directly controlled by the database engine are stored procedures. It is a vulnerable section of code. A stored method returns true or false values depending on whether a client is authorized or not. Attacker will include "; SHUTDOWN; --" with login or secret key for SQLIA. The shown example describes how this type of SQLIA work.

Select username from usertable where user_name='nhj' and pass=''; shutdown;

2. Background Theory –

One of the most dangerous attacks is SQL Injection since it affects the database and has the potential to seriously harm both data and the entire system. SQLIA can have more significant implications than a JS injection or HTML injection, as both are executed on the client-side. In contrast, SQLIA allows you to reach entire database. It is noted that in order to test against this attack, you should have a solid understanding of the SQL programming language as well as a broad understanding of how database queries operate. Additionally, you should exercise greater caution and awareness while carrying out this injection attack because any errors could result in SQL vulnerabilities.

There is much discussion surrounding the use of machine learning algorithms to identify and stop different cyber security risks. The power of using supervised and unsupervised learning approaches to detect security vulnerabilities cannot be questioned, the computer resources and time necessary to execute such sophisticated algorithms remains a key problem for the continually cyber security field. The use of various machine learning methods to identify SQL Injection threats has been the subject of extensive research.

2.1 Machine Learning

Algorithms used in machine learning can learn from data without the need of rules-based programming. According machine learning is a method of allowing a machine to make its own decisions by implementing machine learning algorithm without the use of programmable codes. Machine learning can be divided into two groups. Supervised and Unsupervised.

2.2 Supervised Learning

Supervised learning learns from input data which is labeled data. Machine learns correct answer and makes predictions for future. Labeled inputs and outputs allow the model to measure its accuracy and learn over time. During training, the algorithm can find the relationship between her two variables, so it can predict new outcomes. As a result, supervised learning algorithms can learn better, take on tasks and provide more accurate outputs. Supervised learning divided into two groups.

a. Regression: - Regression is the process of finding correlations between dependent and independent variables. It is useful for predicting continuous variables such as: B. Predicting market trends, predicting real estate prices, etc. The task of a regression algorithm is to find a mapping function that maps an input variable (x) to a continuous output variable (y). Regression algorithms are used to predict continuous values such as price, salary, age.

b. Classification: - Classification is a process that helps divide a dataset into classes based on the similarity of the input dataset. Classification uses a training data set and trains a computer program based on this training. Classify the data into different classes. Classification algorithms are used to classify discrete values such as male or female, true or false, spam or non-spam.

2.3 Unsupervised Learning

This type of machine learning learns from information which is neither classified nor labeled. The machine gives answers without any guidance from the provided information. The machine groups raw data based on patterns, similarities and differences without any training process. This technique

makes prediction without unsupervised. For example, if there are pictures of dogs and cats, the machine cannot categorize whether a picture is a cat or a dog. However, it can categorize them based on the patterns, similarities and differences of dogs and cats.

Clustering is a type of unsupervised method that groups data that are similar to one another and data that are not similar to one another. In unsupervised learning, clustering is the fundamental idea behind collecting data. The intrinsic grouping of the unlabeled data can only be determined using this method. The process of identifying abnormal behaviors is known as anomaly detection. According to Zimek and Schubert (2017), this method identifies unusual items, incidents, or findings that can cause suspicion because they differ significantly from the majority of data.

3. Literature Review –

3.1 SQL Injection Detection System Using Deep Learning [2] - Jothi K.R implements Multilayer Perceptron Model of deep learning to detect SQL injection attack. Neural networks gives a better result and high recall than classic Machine learning model.With the help of MLP model, Author achieved a cross-validated accuracy of 98% with a precision of 98% and recalls of 97%.Furthermore a CNN based architecture can also be integrated to SQL Injection queries for better results. In this paper author demonstrated how websites can implement his machine learning models on their forms and login sessions to confidently detect whether a person is trying to attempt an SQLI query or not.

3.2 A CNN-BiLSTM Based Method for Detection Of SQL Injection Attacks[3]-Neel Gandhi categorized various machine learning algorithm including Logistic regression to decision tree ,ensemble learning technique ,Random Forest ,deep neural networks from convolutional neural networks to Bi-LSTM, and finally propose hybrid model based on CNN-BiLSTM architecture and provide 98% accuracy. Also author presents a comparative study of different types of machine learning algorithms used for the SQLI attack detection. This study shows the performance of various algorithms based on accuracy, precision, search, and F1 score in relation to the proposed CNN BiLSTM model in detecting SQL injection attacks.

3.3 Detection Of SQL Injection Using Machine Learning [6] -Tareek Pattewar uses two machine algorithm Naïve byes and Gradient boosting classifier to detect SQL injection attack. Naive Bayes classifier machine learning model results with an accuracy of 92.8%. Gradient Boosting Classifier learning provide results with better accuracy and selected to be implemented on the SQL Injection classification problem. So Gradient Boosting Classifier from learning is selected to be implemented on the SQL Injection classification problem.

3.4 SQL Injection Detection Using Machine Learning [1]-S.S. Anandha Krishnan analyse different algorithms such as Naïve Bayes, Logistic Regression, Support Vector Machine, Passive regression and CNN for SQL injection attack detection. As result, Support Vector Machine and Passive regression gives 79% accuracy, Logistic Regression and Naïve Bayes provide 92% and 95% accuracy .Highest accuracy provided by CNN of 97%. In this paper CNN is used for future studies and used to create detection model, for better model can use another algorithms and find best model. We will use the static analysis technique here. Runtime analysis can be added as future work. Here only SQL Injection detection system is created but can create a prevention system also in future.

3.5 Machine Learning Predictive Analytics to SQLIA Detection and Prevention [4] - T.P.Latchoumi proposed support vector machine algorithm based model with all possible malicious expression to detect and prevent SQL injection. To overcome SQLI attack, the machine learning concept with the Support Vector Machine (SVM) algorithm is introduced. It is used to detect and prevent SQL injection. In this technique, the SVM algorithm will be trained with all possible malicious expressions and then generate the model. Each time a user enters a new query, SVM is applied to this model to predict whether a particular query contains malicious language. As users invent new techniques, SVM can also detect this malicious expression by matching a minimal set of syntax.

3.6 SQL Injection Attack Detection and Prevention Techniques Using Machine Learning [5] - Ines Jemal provides an overview of the sources, targets, types, and prevention of SQLI attacks. It also provides a comparative table of various techniques that have been proposed to detect SQLI attacks. As a result, after analyzing various proposed techniques, it was concluded that neural networks provide high accuracy. In the future, the author plans to design her machine-learning-based SQLI attack detection scheme, which is lightweight and can handle a large number of requests per second. The authors plan to provide up-to-date standard datasets that researchers can use to evaluate their work and compare with existing solutions.

3.7 Applied machine learning predictive analytics to SQLIA detection and prevention [7]-Solomon Ogbomon Uwagbole presents a complete proof-of-concept for implementing ML predictive analytics and deploying the resulting web service. This web service accurately predicts and prevents his SQLIA using empirical scores presented in Confusion Matrix Receiver Operating Curve(ROC).Confusion matrices provide precision.0.986,Accuracy:0.974,Recall: 0.997. The methodology proposed here works in the context of big data, which is lacking in existing work.

3.8 SQLIA detection and correction using machine learning [8] – Garima Singh propose an algorithm that not only detect SQLIA but also detect unauthorized users by leaving an audit trail log using machine learning clustering. The proposed method provides a model for understanding object behavior. You can verify whether a user is authorized by observing the object's behavior. Audit records are tools for detecting object behavior. A record of ongoing user activity should be maintained and can be used to determine if a transaction is valid. The proposed algorithm leads to a high detection rate. Future work will isolate the techniques implemented as tools compare their effectiveness, efficiency, stability, flexibility and performance and show their strengths and weaknesses.

3.9 SQLIA detection using machine learning algorithm [9] – Joshi Anamika and Geetha V implemented a classifier for SQL Injection Attack detection using a naive Bayesian machine learning algorithm. A Naive Bayes is a probabilistic model that assumes that the values of certain traits are independent of the presence or absence of other traits. The author created a classifier to classify malicious and non-malicious queries. The Naive Bayes algorithm detects attacks based on two probabilities. One is pre-probability and post-probability. Prior probabilities are computed using the specified training set. The number of malicious and non-malicious queries is calculated by calculating previous probabilities. The proposed classifier classifies the test set with an accuracy of 93.3. The proposed method can be improved to detect SQLIA by properly extracting features.

3.10 Machine Learning based technique to detect SQLIA [10]-Muhammad Amirulluqman Azman proposed system to detect an SQL injection attack by comparing the website access log file with the knowledge-based of malicious features. The main methods used are pattern matching and machine learning. Pattern matching method is used to match the log with the pattern which has been made inside the library. Machine learning is another method to use for detection. Machine learning part will undergo some training phase and which will then be used to scan the log for classifying where the log is being injected or not. The classification will result in a malicious or not malicious web request. With the result of 93% in the first test set .It is conclude that the signature-based detection is capable of recognizing for SQL injection attacks.

4. Open Research Issues

To detect SQL injection attacks, there are some issues we discovered by referring to the following research papers. A SQL injection attack allows an attacker to impersonate her identity, manipulate existing data, and cause denial issues such as database servers. High security does not come without cost. Here cost usually means execution time. The first problem is securing existing web applications at low cost. Then back up your stored procedures. The third is detection and validation of malicious SQL code entered by users. Other issues are the methods used to remove and modify malicious input to change it as validated input, reduce execution time, and track input from entry point to SQL statement. It is based on

modeling a dynamic SQL query and comparing it to a set of legitimate SQL query models. These techniques have the advantage of being simple and easy to implement. However, having a model for every legal issue is a lot of work. Without a good query model, a high false positive rate can cause your web application to stop working.

5. Conclusion

SQL injection attacks are a major threat to web applications and can have serious implications for privacy and security. Machine learning applications have been very successful in detecting this kind of web attack. In this review paper, we conducted a systematic literature review of 20 articles dedicated to investigating SQL injection attacks and machine learning techniques. We identified the most commonly used machine learning techniques to detect all kinds of SQL injection attacks. Finally, we present a review paper about analysis of different types of techniques for detecting SQL injection attacks.

6. REFERENCES

[1] S.S. Anandha Krishnan,Adhil N Sabu,Priya P Sajan , A.L. Sreedeep "SQL Injection Detection Using Machine Learning" (2021)

[2] Jothi K R ,Nishant Pandey ,Saravana Balaji B , Abhinandan Amarajan , Pradyumn Beriwal "An Efficient SQL Injection Detection System Using Deep Learning "(2021)

[3] Neel Gandhi , Jay Kumar Patel , Nishant Doshi , Rajdeepsinh Sisodiya , Shakti Mishra "A CNN- BiLSTM based Approach for Detection of SQL Injection Attacks."(2021)

[4] T.P. Latchoumi, Manoj Sahit Reddy, K.Balamurugan "Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention" (2020)

[5] Ines Jemal, Omar Cheikhrouhou, Habib Hamam and Adel Mahfoudhi"SQL Injection Attack Detection and Prevention Techniques Using Machine Learning"(2020)

[6] Tareek Patteewar,Hitesh Patil, Harshada Patil, Neha Patil,Muskan Taneja, Tushar Wadile "Detection of SQL Injection using Machine Learning: A Survey"(2019)

[7] Solomon Ogbomon Uwagbole, William J. Buchanan, Lu Fan"Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention." (IEEE 2017)

[8] Garima Singh,Dev Kant,Unique Gangwar,Akhilesh Pratap Singh, "SQL Injection Detection and Correction Using Machine Learning Techniques " (2015 Springer)

[9] Anamika Joshi,Geetha V "SQL Injection Detection using Machine Learning" (2014)

[10] Muhammad Amirulla Azman, Mohd Fazil Marhusin "Machine Learning-Based Technique to Detect SQL Injection Attack" (2021)

[11] Jamilah M Alkhathami , Sabah M. Alzahrani "Detection Of Sql Injection Attacks Using Machine Learning In Cloud Computing Platform "(2022)

[12] Yi Wang and Zhoujun Li "SQL Injection Detection via Program Tracing and Machine Learning"(2012)

[13] Fitsum Giza chew Deriba, Ayodeji Olalekan SALAU , Shaimaa Hadi Mohammed "Development of a Compressive Framework Using Machine Learning Approaches for SQL Injection Attacks"(2022)

[14] Ogini, P.B , Dr. E.O Taylor, Dr. N.D Nwiabu "A Deep Learning Approach for The Detection of Structured Query Language Injection Vulnerability"(2022)

[15] Zainab S.Alwan, Manal F.Younis "Detection and Prevention of SQL injection Attack"(2017)

[16] Ashwat Keshav Hedge ,P N Jayanthi "A Survey on SQL injection Attacks and Prevention methods" (2020)

[17] Premveer, Ankur Srivastava,Anurag Jain "Vulnerability detection for SQL injection attack" (2013)

[18] Maha Alghawazi ,Daniyal Alghawazi "Detection of SQL injection attack using machine learning techniques "(2022)

[19] Pushpendra Kumar , R.K. Pateriya "A survey on SQL injection attacks, detection and prevention techniques"(2012)

[20] Jianwei Hu ,Wei zhao "A Survey on SQL injection Attacks, Detection and Prevention"(2020)