An Efficient SQL Injection Detection System Using Deep Learning

Jothi K R
School of Computer Science and Engg
VIT university
Vellore, India
profjkr@gmail.com

Saravana Balaji B
Dept of Information Technology
Lebanese French University
Erbil, Iraq
saravanabalaji.b@gmail.com

Pradyumma Beriwal
School of Computer Science and Engg
VIT university
Vellore, India
pradyummaBeriwal1999@gmail.com

Abhinandan Amarajan
School of Computer Science and Engg
VIT university
Vellore, India
abhinandan98579@gmail.com

Nishant Pandey
School of Computer Science and Engg
VIT University
Vellore, India
npandey039@gmail.com

Abstract—SQL Injection makes most of the applications that are based on different types of databases be it used in any devices vulnerable to cyber threat. SQL Injection is said to be one of the top most threat that database-based applications on the web. SQL Injection makes all the user’s information present in the database vulnerable and the user’s data may be either sold in black market or may be misused. The disadvantages of previously implemented SQL model is that they will not know how will they be able to categorize new patterns, they will only be able to detect the patterns which they have experienced before or trained on. But our model will be able to identify whether the data entered is SQL injected or not identifying patterns in the input. The advantages to our system will be that it will be able to detect all and every type of Injection techniques. All the feature extraction and selection will be done by the model itself. Just the user should need to enter the text. It is also scalable and can extend it to a wide variety of applications. With the help of MLP model, we have achieved a cross-validated accuracy of 98% with a precision of 98% and recall of 97%.

Keywords—SQL Injection Detection; Deep Learning; F1 Score; Deep Learning; Types of SQL Injection

I. INTRODUCTION

In Today’s time there are billions and billions of web-based applications (examples may include E-Commerce, Chatbot, Social Media ‘s, banking services in fact each and every service) that are dependent on some sort of a database. Those databases may be either relational (SQL) or non-relational (Mongo DB, Cassandra DB). Most of them are vulnerable to cyber threats and the most common one that they face is SQL Injection [1]. The SQL Injection has been existing for a large number of decades, and the damages caused by then has also been increasing exponentially. Now due to great connectivity across the world and availability of resources easily companies are more prone towards cyber-attacks [2]. Companies have been selling millions of dollars in order to find a proper solution for SQL injection attacks [3]. There are 3 basic components in the system- Authentic User, Detection Module, and Malicious User.

II. SYSTEM SCOPE

Authentic User- An Authentic User uses the website for his normal work and services. He uses a valid Email id and Password to access the website.

Detection Module- The detection module correctly classifies any attempt to use SQL injection on the application and block that user.

Malicious User - The malicious user can use SQL injection methods/commands to get access to the application.

Using Machine Learning and Deep Learning in order to find a solution to prevent cyber-attack has been reduced significantly [4]. One down sight to Machine Learning is that it is not able to recognize if new patterns are formed. But using deep learning, this can be overcome.

III. SQL INJECTION TYPES

Error Based: here the attackers will perform the injection which will indirectly produce error messages and get information about the data in the database. They generally use it to get the structure of the data. Union Based: here the attackers use the union operator in order to combine SQL queries in order to extract, delete or modify data in the database. Blind: It will take some extra time for the attacker to make an attack. Here the attack won’t receive the data directly from a web application. Rather he will be sending payloads of data to the server and will try to get the information about the structure of the database [5][6].

IV. WORK COMPARISON WITH REFERENCES

Major Issues found in references: Inability of the model towards some types of SQL injection attacks [7]. Some models depend on a fixed algorithm and unable to detect new patterns [8]. Some models are using classical Machine Learning Models which may be poor when compared to a deep learning model [9][10].

Solutions in the proposed system: Our model has been trained on a dataset with many instances of all types of SQL attacks. The high accuracy supports the models ability to detect all types of attacks. A deep learning model with the embedding layer technique has been used which depends upon raw data instead of pre-defined features which prove out to be ineffective.

Disadvantages of the current systems include that the model relies on the single word tokenization and does not incorporate for N-gram models. This makes the model unable to get context of the complete sentences. This may