Detecting The Normal And Injected Sql Query Using Random Forest Classification With Rabin Carp Pattern Matching Algorithm In Web Database

Poonam Acharya[1], Meeta Sharma[2]

[1]Research Scholarity, Government women engineering college Ajmer
[2]Assistant Professor, Government women engineering college Ajmer
[1] poonam.acharya24@gmail.com ,[2]Meetasharma@gweca.ac.in

Abstract: Nowadays detection of SQL injection attack is becoming very important in database-driven websites. SQL-Injection attack is a class of attacks that many of these systems are highly vulnerable to, and there is no known fool-proof defense against various attacks. In order to overcome this problem, the assurance for security to all the web Sites is essential for the protections for their database. Various Companies facing challenges in intrusion and Detection while installation and deploying their sites. And most common attack in the way if requesting the input as malicious logic in sql query of the authentication page. So that query returns success response in the condition of functionality. In this paper, we proposed the two algorithms such as Random forest and Rabin carp pattern matching to detect the injected query attack. Random forest is the technique used to detect the anomaly queries from the dataset. And Rabin carp pattern matching is used to prevent SQL injection in authentication page.

Keywords: SQL injection, Random forest, Rabin carp

1. Introduction

Human beings are a creature that can think differently from other living things and question both their own actions and their environment. With this feature, he tries to learn his surroundings by showing tendencies such as discovering, researching, asking questions, and noticing the relationships between objects. In other words, he tends to understand the world he lives in with various judgments. In this process, these judgments developed by the individual affect his life significantly as well as his education life deeply. In parallel with this, information and communication technologies have covered every stage of human life. (Kumandaş & Kutlu, 2014; Tiemann, 2011) The temporal difference between generations has gradually narrowed. The world, lifestyle and social change seen by a child born in the nineties and a child born in the 2000s are quite different from each other. Therefore, education cannot be done with the methods, tools and teaching techniques of the nineties or earlier. This situation has caused changes in education and training programs in our country as well as throughout the world. Therefore, this situation has made a new perspective necessary in mathematics teaching. "New knowledge, opportunities and tools are reshaping our perspective on mathematics, our expectations from mathematics, the way we use mathematics, and above all our mathematical learning and teaching processes." (MEB, 2013). This new perspective has made it a priority target to raise individuals with advanced reasoning skills, mathematical thinking skills, proofing skills, problem solving skills, metacognitive knowledge, skills or qualifications. The realization of these goals can only be achieved with an education that is suitable for these skills. Goswami (2004) stated in his study that children obtained true or false conclusions and inferences by conducting inductive or deductive judgments in various aspects from the first years of their lives. NCTM (2000) also stated that children make very different judgments in primary education and in the following years, and they can make various inferences, right or wrong, according to their own assumptions. Therefore, it can be thought that the strategies developed by children at this age will be determinant in their future education lives. Because reasoning is one of the basic concepts at the center of teaching and learning mathematics, as in all areas of life (Russell, 1999). However, one of the concepts that opens the door to reasoning is to increase the understanding capacity of the learning individuals. For this purpose, special lectures or courses are given to increase understanding skills in developed western societies in the field of education such as the USA (Francisco & Maher, 2005; Generazzo, 2011; Hiebert & Grouws, 2007; Hsu, 2010; Lee, 1999; Martin & McCrone, 2009; Pulley, 2010). Thus, in terms of conceptual continuity, concepts that form the basis of mathematics such as mathematical thinking, association, abstract thinking, intuition, and communication can only be acquired by teaching correct reasoning. It can be said that this can only be possible with the right teaching models, methods, techniques and teachers who can use them in the most efficient way.

Designing educational programs; It is important to determine appropriate models, methods and techniques in the realization of teaching. Shulman (1986) emphasized the importance of the relationship between pedagogical content knowledge and field knowledge, and drew attention to the technique used and its practitioner. In other words, teachers play an important role in addition to using appropriate teaching methods and techniques to overcome the difficulties encountered in mathematics teaching and to enable students to make correct judgments. Therefore, it is very important that teachers who contribute to the realization of teaching use appropriate teaching models, methods and techniques in this process. In this regard, teachers; It is of great importance for students to be aware of reasoning processes, to make instructional decisions in order for them to develop their reasoning with correct strategies and to minimize their mistakes (Güven & Demir, 2015). Therefore, teachers who are at the center

of teaching; Knowing the mathematical and mental thinking processes of students and giving them the necessary feedback for this requires them to be aware of possible reasoning errors (Petrou & Goulding, 2011). This situation is directly related to the education teachers receive and therefore their knowledge of teaching mathematics. Because, if a teacher's understanding and infrastructure regarding the acquisitions and concepts to be taught is limited, it may prevent the teaching process from being carried out efficiently. Therefore, it can cause students to make many reasoning mistakes in the learning process or to continue their reasoning mistakes (Stylianides & Stylianides, 2009). In this sense, teachers must first be equipped themselves in order to reveal students' skills such as reasoning, problem solving, abstraction, and proof (Baki, 2008). Otherwise, develop students' correct judgments; It may become more difficult for them to develop their skills in understanding, solving and interpreting mathematical problems (İskenderoğlu & Baki, 2011).

Teachers' understanding of the basic components of modern mathematics teaching such as mathematical thinking, proof, reasoning and problem solving, the strategies they put forward and the methods they use have a great influence on the shaping of this process. In other words, if a teacher does not have sufficient knowledge to develop reasoning strategies and reasoning in the learning-teaching process, it cannot be expected that his students will gain this understanding or these skills. For this, it is important to know the understanding of teachers towards the teaching process of mathematics learning, the methods they use in this process, the strategies they put forward, or how they evaluate the process (Healy & Hoyles, 2000). NCTM (2000) also reveals the importance of teachers in the process by stating that teachers' field knowledge and understanding, knowledge about students and teaching strategies contribute to the effectiveness of mathematics teaching. Therefore, it is important for teachers to know the reasoning they put forward while dealing with the problems that their students have solved, the problem-solving steps they have followed in this process, the reasons they have presented for these steps, to recognize the relationships between the stages and to evaluate them correctly. This situation reveals the importance of teacher training programs. Because it is important that the mathematics teaching process is carried out properly and the actions inherent in basic skills are kept alive, but the role of teachers in this process cannot be denied. In this process, teachers play an important role in helping students gain skills such as making sense of mathematical concepts and the relationships between these concepts, interpreting and associating them with other mathematical concepts and topics (Kinach, 2002; Vale, McAndrew & Krishnan, 2011; NCTM, 2000; Van de Walle, 2013). Thus, it is thought that students who gain this skill will learn mathematics more meaningfully, value mathematics and permanent learning will be realized (Businskas, 2008; MEB, 2013).

2. SQL Injection

Basically what do we know about SQL injection is that whenever an attacker wants to update or modify the information in the database that can be done by the SQL injection itself. Assume a webapp creates the accompanying SQL articulation:

SELECT writer, title, year FROM books WHERE distributer = 'BPB' and published=1

In the event that an attacker to enter a string, for example,

BPB' OR 1=1—

into the hunt structure, this would bring about the accompanying question:

SELECT writer, title, year FROM books WHERE distributer = 'BPB' OR 1=1—' and published=1

This inquiry would return each row in the data set where the distributer is BPB, or 1=1, which is in every case valid, therefore returning all the rows in the database [23].

There are a lot of existing techinque that can analyse about the SQL attacks. One method being used includes static examination of code to frame a model that can measure up to dynamic inquiries at runtime to distinguish SQL infusion attacks. In Lee et al. [6] the creators propose a framework that performs investigation of PHP-based sites to assess the SQL questions, at that point measures these inquiries by eliminating boundaries to shape general inquiry structures, a cycle they allude to as question change. These are then contrasted with live SQL questions created powerfully and an inquiry that has an alternate construction will be hailed as an attack. As well as eliminating boundaries, Kar et al. [8] sum up SQL questions into underlying components, and had the option to accomplish a 100% identification rate with their method.

There are very different design had been investigated for Intrusion Detection Systems (IDS). In Djanali et al. [5], the users make bunched design that can be processing by Blackbeerry processing gadgets. Here the task utilizes the High Interaction Honeypot Analysis Toolkit (HIHAT) HIHAT honeypot framework [10] and also Structured Query Language infusion identification strategy proposed in Lee et al. [8]. The framework utilizes a heap adjusting worker to course calculation to the Blackbeerry processing gadgets.

K OKILAVANI, T et al. [7] introduce an appropriated, multi-honeypot design. The author utilize a few distinctive honeypot frameworks associated from a front-side host. The author use the framework is secluded for oblige the expansion or evacuation of multi-honeypot frameworks. There is a dispersed design has the benefit that it's equipped for social occasion more information than any of the honeypot frameworks, and the creators focus on the attacks they distinguished were fundamentally focusing on SSH, just as MySQL, MSSQL, and telnet, and begin basically in Russia, China and the European Countries.

P ARSA et. al [8] introduce a convoluted engineering they allude to as a "virtual honeynet" intended to amplify the period that has spent by an assailant in the framework and hence the measure of information accumulated about aggressors. This framework utilizes the SNORT [19] IDS to distinguish noxious traffic which is then redirected to the honeynet. The virtual honeypots utilize the HIHAT [9] framework just as the Sebek [30] information catch instrument as parts, and consolidate these into a convoluted engineering intended to befuddle and defer assailants.

In the Honeydoop [9] framework, the creators introduce a powerful engineering utilizing Hadoop for information assortment. Here the framework utilizes frequent allotment dependent on as of now noticed organization conditions, with the objective of proficient usage of assets and expanded security. The creators notice that as far as they can tell these powerfully made honeypots get considerably found much traffic pattern than produced the fixed. The creators use R AHMAN [11] for traffic catch, yet the engineering is particular so any IDS framework could be utilized

3. Architecture Diagram
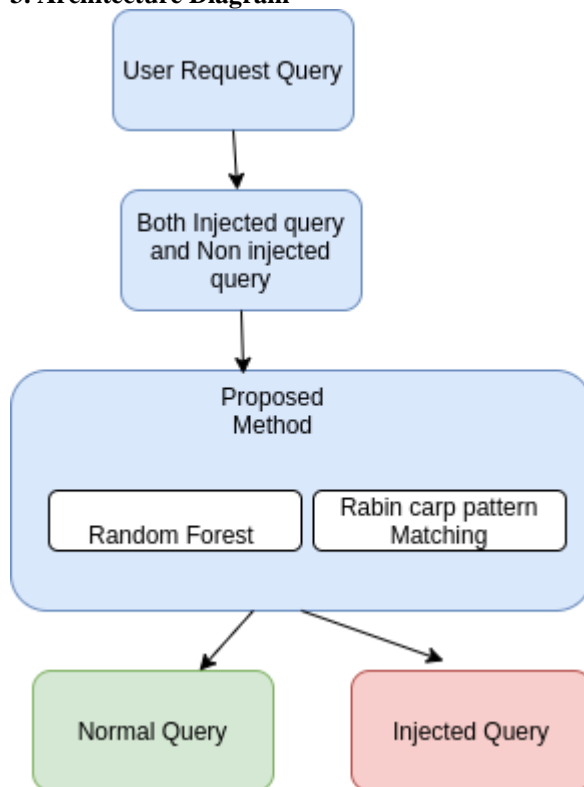
Our proposed method is used to detect and prevent the SQL injection. Also, this works classified with two parts such as our first part to be classification of normal and injected query. Classification is works based on the dataset such as injected and non-injected query sets. The malicious logic or code requested to database. This observation captures track the data is considered as SQL injected query. The Second part is preventing the SQL injection using Rabin carp pattern matching algorithm in authentication page of web pages.

Step by step procedure of the process
Step 1 - Training set selection
Step 2 - At the initial condition system input is SQL-Query string.
Step 3 – In order to generate the model inside the random forest process the training set is feed.
Step 4 – The next process is prediction
Step 5 – Followed by that the classifier is used to classify the model.
Step 6 – The accuracy of the algorithm is calculated by using the labelled outputs.
Step 7 – This process is repeated until to reach correct classification precision.

3.1 Random Forest Algorithm:

From the earlier study we analysed that Leo Breiman and Adele Cutler created this algorithm and it is compared with "Bootstrap aggregating" methodology as well as "random subspace methodology". It helps to create decision tree which leads to help for the process of classification.

There are many dissimilarity is present between Random Forest and the decision tree algorithm. One among the major dissimilarity is root node finding and feature nodes splitting in the random manner.

Inside the random forest algorithm several decision tree classifiers are present and the mode of decision tree classification grades are responsible for the output categories.

Two random selection process of random forest algorithm are preceded during the construction of a single decision tree. Those selection processes are training samples random selection and by this algorithm we classify the Normal and SQL Injection query in the web pages.

Pseudo code for Random Forest Algorithm:
1. At the initial condition "S" features are randomly chosen from the total features "t" where S<<t.
2. In the middle of the "S" features, by the use of the finest split point the node "n" is calculated.
3. The nodes are split into child nodes by using the finest split point.
4. The steps 1 to 3 are repeated till to reach the number of nodes "I".
5. The forest model is built by repeating the steps 1 to 4 for "T" times to create the number of trees.

Secondly after the formation of the random forest classifier, prediction is carried out. So the pseudo code for random forest prediction is given below:
•   The tests features of random forest decision model are taken randomly as well as the outputs are predicted using the randomly created decision tree and the predicted outputs are collected then stored.
•   Calculation of Predicted target votes.
•   Among the Predicted target votes the highly voted targets are considered as the final prediction value.

3.2 Rabin carp pattern matching algorithm:
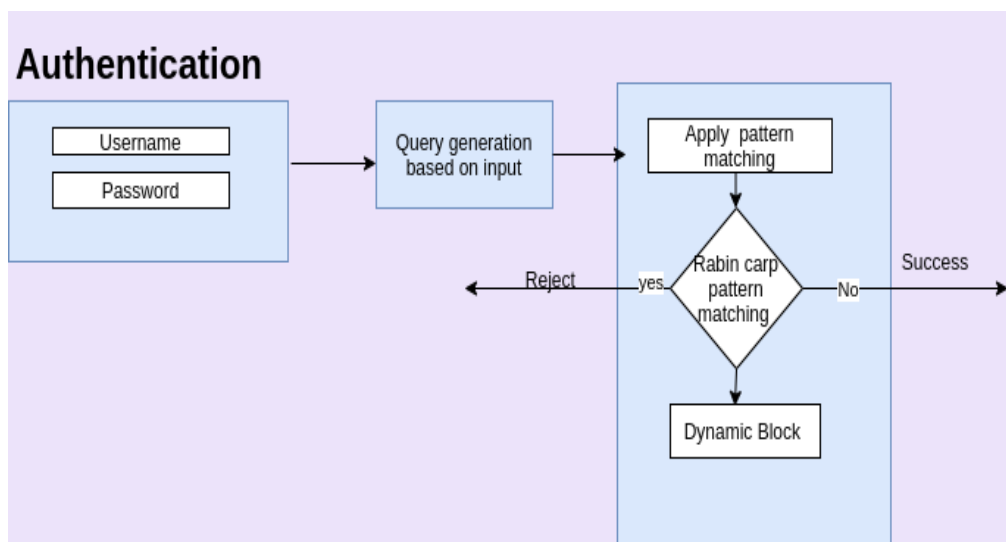
In Login or Authentication process, User needs to give the username and password. By using the information system creating the query for authorization. Pattern matching algorithm compares the query with static pattern list. We need to maintain the list of some anomaly pattern. Dynamically blocks with IP address if matching algorithm returns true.

3.3 Classification of query with Rabin carp pattern matching algorithm

Step 1 – Selection of new query * from Admin whereas userID = "1"

Step 2 – Selection of mistrustful query * from Admin whereas userID=" "OR 1=1;--"

Step 3 – The new query is transmitted as well as the mistrustful query is jammed.

Here SQL-query strings maintain a work list to store the tokens.
"N" denotes new query and "M" denoted mistrustful query
Step 4 - Selection of ("N") * from Admin whereas userID = "1"
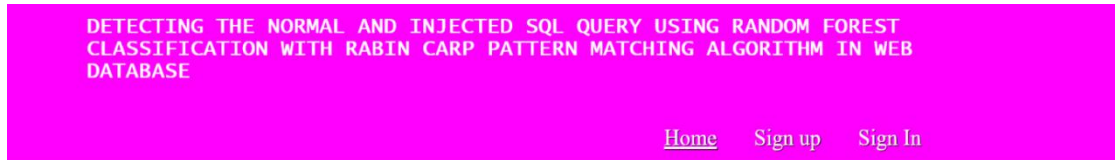Step 5 - Selection of ("M") * from Admin whereas userID = " " OR 1=1;--"
Step 6 - Selection of ("N") * from Admin whereas userID = "1" && password = "abc";
Step 7 – Selection of ("M") * from Admin whereas userID = " " OR 1=1;--,

4. Implementation and   Results
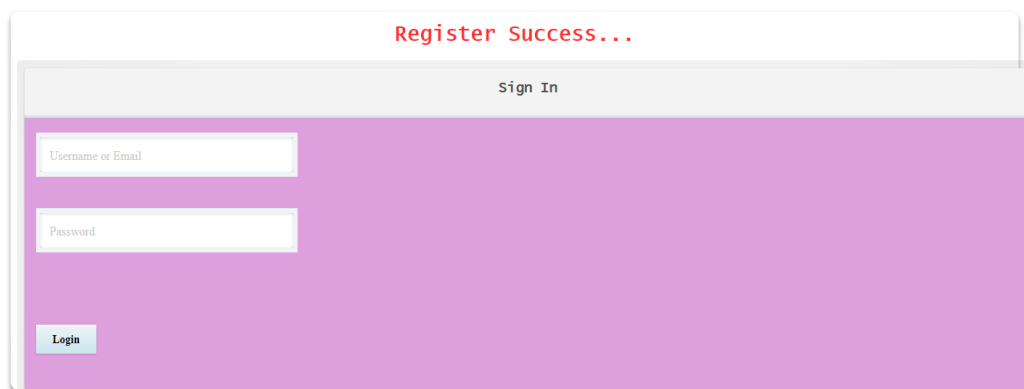
SCREEN SHOTS
### SCREEN 1: Home page

SCREEN 2: Registration page

User needs to register the application before login. So that authenticated user login the application. The fields are Username, Password, Email ID and Number.

SCREEN 3: Registration Success

Register successfully done with right credentials. If not registered successfully the error message throwing on the web page.

SCREEN 4: Login Screen:
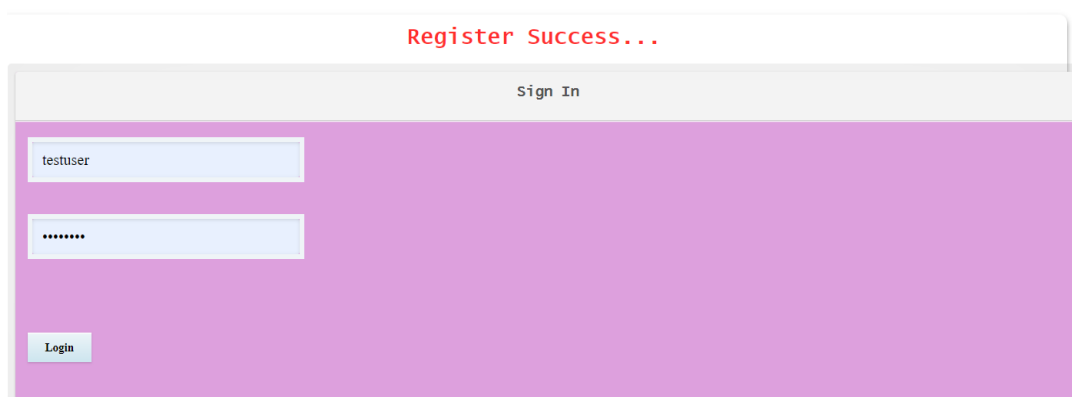
User needs to login with username or email and password to access the application. If user have to enter the correct username or password. If error message will throw as Invalid Username or Password

SCREEN 5: Classification of SQL Injection

Based on our proposed system, SQL injected are classified by Random forest algorithm. The dataset contained SQL Injected and normal queries. Based algorithm RF algorithm dataset are classified with 96 percent accuracy.

SCREEN 6: SQL Query injected in Authentication page

Any attacker requests the malicious code or logic by the way of SQL query like
Example:
select * from registration where username= "or ""=" and password= " or ""=".
So query passing the condition and will goes to next page of web screen.

SCREEN 7: Dynamic block by Rabin carp pattern matching

Login Blocked and stored in database with IP to identify the malicious entry of unknown data by the users Address because of attacker requested the injected query.

.
5. Conclusion

In this work, we have proposed a system for detection and prevention of SQL injection using the Random forest and Rabin carp pattern matching algorithm. Random forest is the technique used to detect the anomaly queries from the dataset. And Rabin carp pattern matching is used to prevent SQL injection in authentication page and blocked the attacker with IP address to identify the malicious entry of unknown data by the attackers.

References

1. A BELS, T., DHAWAN, P., and CHANDRASEKARAN, B. (2005), "An overview of xen virtualization," Dell Power Solutions, vol. 8, pp. 109–111, .
2. Abhishek Kumar, Pramod Singh Rathore, Vishal Dutt,(2019) "An IOT Method for Reducing Classification Error In Face Recognition With the Commuted Concept Of Conventional Algorithm", International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019, ISSN: 2278-3075, 8(11), DOI: 10.35940/ijitee.J9861.0981119.
3. B ITTENCOURT, L. F., MADEIRA, E. R., CICERRE, F., and BUZATO, L(2005)., "A path clustering heuristic for scheduling task graphs onto a grid," in 3rd International Workshop on Middleware for Grid Computing (MGC05), .
4. BITTENCOURT, L. F., SENNA, C. R., and MADEIRA, E. R.(2010), "Scheduling ser-vice workflows for cost optimization in hybrid clouds," in Network and Service Management (CNSM), 2010 International Conference on, pp. 394–397, IEEE,
5. Bhargava, N., Bhargava, R., Rathore, P. S., & Kumar, A. (2020). Texture Recognition Using Gabor Filter for Extracting Feature Vectors With the Regression Mining Algorithm. International Journal of Risk and Contingency Management (IJRCM), 9(3), 31-44. doi:10.4018/IJRCM.2020070103
6. C HAWLA, Y. and BHONSLE, M.(2012), "A study on scheduling methods in cloud com- puting," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 1, no. 3, pp. 12–17, 2012.
7. E LZEKI, O., RESHAD, M., and ELSOUD, M., "Improved max-min algorithm in cloud computing," International Journal of Computer Applications, vol. 50, no. 12, pp. 22–27,
8. G RAY, M.(2010), "Cloud computing: Demystifying iaas, paas and saas," Retrieved July, vol. 17, p. 2011,.
9. K OKILAVANI, T. and AMALARETHINAM, D. D. G.,(2011) "Load balanced min-min algorithm for static meta-task scheduling in grid computing," International Jour- nal of Computer Applications, vol. 20, no. 2, pp. 43–49, 2011.
10. L IU, G., LI, J., and XU, J.(2013), "An improved min-min algorithm in cloud comput- ing,"   in Proceedings of the 2012 International Conference of Modern Computer Science and Applications, pp. 47–52, Springer,.
11. M ELL, P. and GRANCE, T (2011)., "The nist definition of cloud computing,".

12. P ARSA, S. and ENTEZARI-MALEKI, R.,(2009) "Rasa: A new task scheduling algo- rithm in grid environment," World Applied sciences journal, vol. 7, pp. 152–160, 2009.
13. Rathore, P.S., Chatterjee, J.M., Kumar, A. et al (2021). Energy-efficient cluster head selection through relay approach for WSN. J Supercomputing. https://doi.org/10.1007/s11227-020-03593-4
14. R AHMAN, M., LI, X., and PALIT, H., (2011)"Hybrid heuristic for scheduling data analytics workflow applications in hybrid cloud environment," in Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE Inter- national Symposium on, pp. 966–974, IEEE, 2011.
15. V AN DEN BOSSCHE, R., VANMECHELEN, K., and BROECKHOVE, J.,(2010) "Costoptimal scheduling in hybrid iaas clouds for deadline constrained workloads," in Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on, pp. 228–235, IEEE, .
16. V AN DEN BOSSCHE, R., VANMECHELEN, K., and BROECKHOVE, J.(2011), "Cost efficient scheduling heuristics for deadline constrained workloads on hybrid clouds," in Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on, pp. 320–327, IEEE, 2011