Open access Journal International Journal of Emerging Trends in Science and Technology

SQL-Injection Vulnerability Analysis Using Machine Learning Technique

Authors
Gunjan Shukla[1], Professor C.S. Satsangi[2]
[1,2]Information Technology, Medicaps Institute of Science and Technology, Indore, India
Email: [1]*gunjan.shukla1@gmail.com*, [2]*cssatsangi108@gmail.com*

Abstract
Internet may be the supreme supply of data, data is situated in distinct data format and can be accessible everywhere you go, consumer will get connection together with web by way of request covering available as GUI screen, signifies Net is accessible through the use of browser in which consumer might feed his/her data pertaining to authentication in case required through request. Seeing that on view natural environment involving web connection and design development, distinct authentication mechanism, security password safety and an incredible number of protection treatments have been made to defend the approval via unauthorized entry but still crooks are aimed towards distinct ways to break the actual protection, it may be through hit and walk methods, through infecting computer system, through surging computer system, But within the actual suggest cardstock a fresh strategy have been offered to get SQL-Injection being exposed, in case offered in user's suggestions, the item assessments dilemma personal, finger prints and mapping blend to help think any intruding activities throughout the process,
This suggest strategy is simple to use, since it simply desires fingerprinting and mapping paradigm involving dilemma and all too easy to change, in case new personal is found, instead of positioning any overhead within the existing doing work process.
Keywords— *SQL-Injection; SVM; Attack;*

I. INTRODUCTION

This Injection is the attack which occurs on application layer and only needs database fingerprinting and flaws in designing process. It mostly occurs due to improper conduction of back-end queries, Injection or poisoning has been classified in different types like SQL-Injection, Script Injection, Shell Injection, XML Injection, HTML Injection[1] etc. It is impossible to protect system for life by creating safe system design once, It must be updated by regular interval of time, as technology is getting advanced multi ways are available to track user, hacker are well trained and behavior based internet-thief ,it takes advantage of mistakes done by either designer or may be by users.

So for detecting and blocking the attack a simulation tool had been designed, which will detect web attacks and also block them, and if in future new signature is found could be easily updated in the system. Different testing procedure are there to test multiple cases but, what is the system is not designed for them ex: system is designed to detect attack[2] of particular type it could not be able to detect attack of another type until it is not updated for it.

Attack is hit and trial methods accomplished by several methods, several parameters [3] affects security environment like open input enviornm-nnt, number of attempts user gets achieve access, vulnerable environment due to insecure design and insecure parameter assignment and declaration.DOS Attack, Phishing attack, and several more attacks ,which are carried out by sending continuous malfunction request to site for confusing server and flooding them, due to which server and system gets busy in solving and calculating their authenticity and creates passing or processing of unauthorized query so as to

perform intruder activities, And phishing attack is carried by creating or changing web url or login credentials ,it also carries url input ID to cause or change their values ID, here the purpose of attacker is to execute query at back end. Attacker also creates dummy page[4,5],which looks exactly like original page but when user feeds data in the page it diverts access credentials towards suspicious page, where it saves information like username and password and query processing information to hack user directory page.

Attack could be Framed at application layer by providing unwanted information or input, and other attack[6] based on network layer ,targets attack by routing methods ,where it routes packet to untargeted way and thus forges the original data to intruder. here the intruder diverts the actual traffic to other routes and misguides the system by creating suspicious id and time delay as it needs to add other framing fields in routing system[7].

SVM (Support Vector Machine)

SVM is machine learning based classification techniques which classifies data based on support vectors, it is based described for the classification of data found in two classes, SVM [24] linearly separates data in two separate hyperplanes. Consider the example shown in Fig 1. Here two classes are classified in hyper plane, thus separating the support vector without losing the originality of domains. SVM work on datasets of training data and testing data described by ATTIBUTES and LABELS means their classes and by their features. It is a well defined modeling techniques for predicting and classifying the values found in classes. It separates date available near margins, called as supports vectors. It is best suited for classification but only restricted to two class problems only.

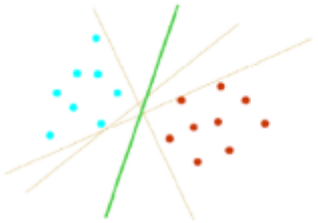


Fig1. Optimal Separating Hyper Plane

Based on the dataset of training data and testing data 4 different functions are defined, which are mentioned below, they are best suited for data available in higher dimensional space for classification.

To attain this goal there are four different kernel functions.

1. Linear: $K(x_i, x_j) = x_i^T x_j$
2. Polynomial: The polynomial kernel of degree d is of the form.

$$K(x_i, x_j) = (x_i x_j)$$

3. RBF: The Gaussian kernel, known also as the radial basis function, is of the form

$$K(X_i, X_j) = exp\left(-\frac{\|(x_i, x_j)\|}{2\sigma^2}\right)$$

4. Sigmoid: The sigmoid kernel is of the form

$$K(x_i, x_j) = tanh(k(x_i x_j) + r)$$

Here Propose work is implementing Linear Kernel function techniques, which is a special case of RBF (Radial Basis Function). RBF separates nonlinear data it hyper plane or higher dimensional space.

II. RELATED WORK

Static analysis framework (called SAFELI) has been proposed by X iang Fu et al [5], for recognizing SIA (SQL Injection attacks) weaknesses at compile time the source code information are often evaluated by SAFELI and can be able to recognize terribly delicate weaknesses that cannot be exposed by black-box vulnerability scanners.

The mechanism to keep track of the positive taints and negative taints is recommended by William G.J. Halfond, Alessandro Orso, Panagiotis Manolios [10], Defensive Programming [11] [12] has given an approach for Programmers by that they will implement their own input filters or use

existing safe API s that prevent malicious input or that convert malicious input in to safer input. Unprotectedness pattern methodology is employed by Livshits and Lam [8], they propose static analysis approach for locating the SQL injection attack. The most problems with this methodology, is that it cannot discover the SQL injection attacks patterns that do not seem to be known beforehand. Vulnerability patterns are delineate here during this approach.

Automated attack detection tool Qualys [8,9] has been designed that detects with their signature and fingerprints and additionally generates report for code correction at nominal position.

Different Static analysis tools has been designed named IDs [10,11], they discover attack on the parameters provided to them, they traces their behavior and signatures referred to as kali linux.

D-Word [13,15] methods had been designed to detect DDos attack, which detects and stops unwanted request appearing from other system on network, it is flexible and detects parameters and time delays to detect destined attacks.

Vulnerability scanning of a network needs to be done from both within the network as well as without (from both "sides" of the firewall). The approach I would suggest is to start from the network evaluation phase [10,12], where sniffing and primary attacks are performed. The gathered data is used in the attack phase to exploit the exposed vulnerabilities.

Wireshark tool designed to filter attacks appearing by TCP broadcast, it is GUID [17,18] based system and detects spoofed packets and, to capture communication between two IP addresses, or capture UDP-based [22,23] DNS queries on the network. Traffic data can be dumped into a capture file, which can be reviewed later. Additional filters can also be set during the review.

III. PROPOSE WORK AND EXPERIMENTAL APPROCH

In the propose work a unique technique on web application attack detection has been implemented to detection suspicious and malicious activities, a system has been designed here to detect signature and finger prints of web attacks, which is based on machine learning techniques, here support vector machine(SVM) is used for classification of attacks ,based on classes defined below Original class or safe or authentic class represented by "O" Malicious class or unsafe or suspicious class represented by "S" A dataset has been designed containing signatures of both the class that is of safe class and unsafe class.

A. *Example:* ***SQL Query Dataset***

Select * from college where uname="abc"; **Safe Class (O)**

Select * from college where uname " "OR 1=1 ; **Unsafe Class (S)**

B. *Example:* ***Socket Dataset (IP address+ Port Number)***

342.12.23.167	**Safe Class (O)**
196.345.12.43.22	**Unsafe Class (S)**

C. *Example:* ***URL Keywords Dataset***

www gmail com	**Safe Class (O)**
www gmeil uk	**Unsafe Class (S)**

Algorithm

1. *Select reasonable amount of dataset for training:*
   a) *Select SQL Query Dataset*
   b) *Socket Dataset*
   c) *URL Keywords Dataset*
2. *Take input from the user*
3. *Check Input data.*
   a) *If input data matched with Suspicious class query will blocked*
   b) *If matched with Original Class Query will pass.*
4. *Calculate Different parameters (Detection Time, Training Time, TPR, TNR, FPR, FNR and Accuracy) based on query fired.*
5. *Process the system for different dataset size, taken to calculate different parameter values.*
6. *Repeat the steps 1 to step 5 until proper precision is not achieved.*

FLOW CHART

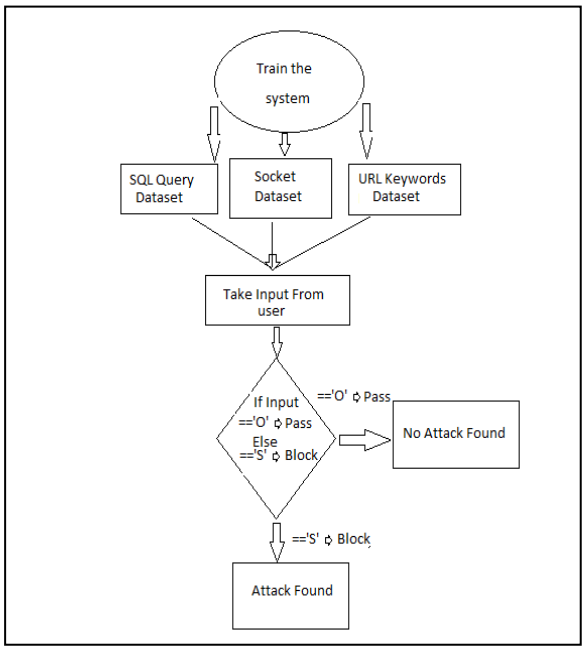
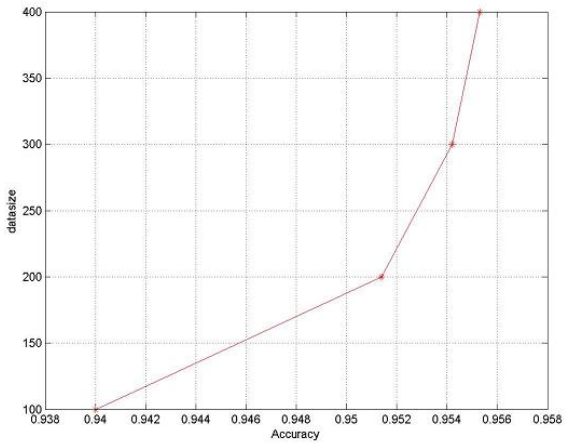


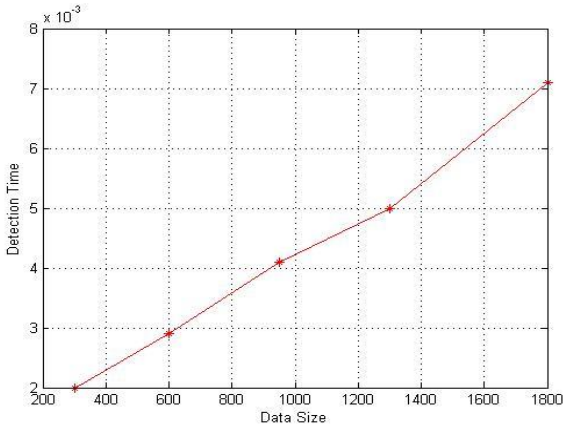
Fig. 1. Flow Chart for algorithm

IV. RESULT ANALYSIS

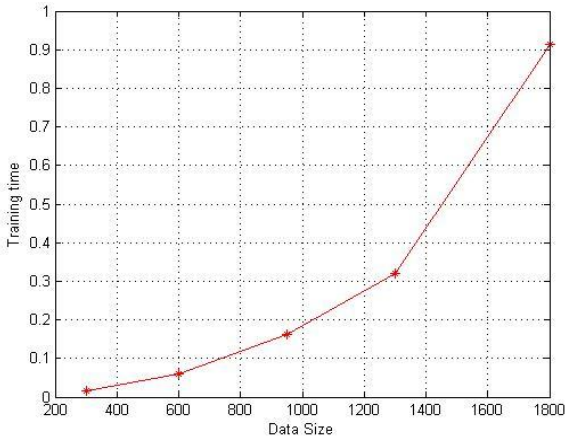
The propose framework has been tested on dataset taken in different quantity ,and different parameters has been calculated based on the behavior of dataset size ,detection time, training time, accuracy. The observation is created on different parameters taken 2 at a time and their fluctuation has been calculated. As dataset of different size has been taken ,it is found that the Accuracy is 96.3% which is best among the available techniques ,as it is a light weight system, means easy to configures ,easy to modify ,if new attack signature and finger prints are discovered.



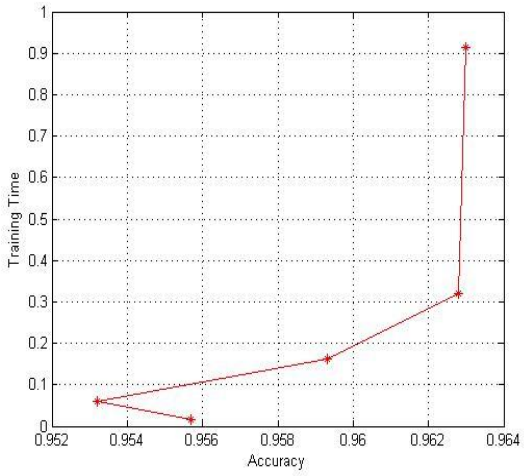
The dataset size of different interval is taken and accuracy is calculated and it is found system is showing nearby linear behavior.



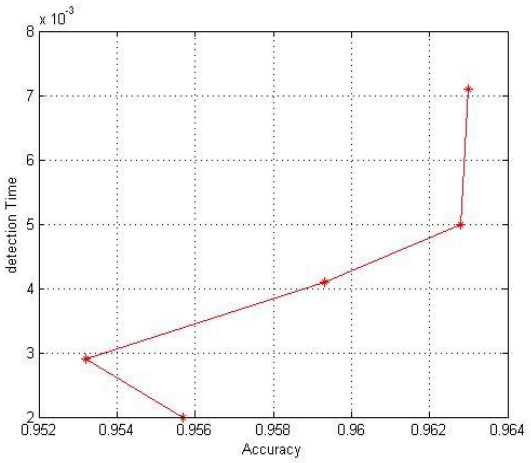
Here, it is clearing predicting that by increasing the dataset size, detection time is also increasing, means the growth is linear.



The Training time increasing when dataset size increases, as it based on features based system.



The training time is continuously varying at regular interval when training time varies.



The training time is continuously varying at regular fixed interval when Detection time varies.

V. FUTURE WORK

The Propose system has been only defined for SQL Query, Socket Query, URL Keywords types of attacks, but in future could be enhanced to track all multiple attacks by which web application layer is susceptible too. As web application is directly accessible by millions of users, means millions of attacks are possible, which are falling in different classes of attacks.

VI. CONCLUSION

The propose system has been calculating different marked web application attacks and classifying their types based on class provided to it, the system is convenient and light weight and easy to implement on existing system, Sit is taking little overhead and also flexible to add new signature of dataset ,if found on the system, the current system is only checking SQL Query, Socket Query, URL Keywords, The system has been tested on different dataset size and it has provided accuracy of more than 96.3% ,which is found to be best among the all available system.

Acknowledgment

The Write as per your requirement.

REFERENCES

1. Alexander Schaub,Emmanuel Schneider, Alexandros Hollender ,"Attacking Suggest Boxes in Web Applications Over HTTPS Using Side-Channel Stochastic Algorithms", CRiSIS 2014 conference
2. A Classification of SQL Inject ion Attacks and Countermeasures: William G.J. Hal Fond and Alessandro Orso, Col lege of Computing, Georgia Institute of Technology.Gatech.edu.
3. D. Scott and R. Sharp, "Abstracting Application - level Web Security", In Proceedings of the 11th International Conference on the World Wide Web (WWW 2002), Pages 396–407, 2002.Y. Huang, F. Yu, C. Hang, C. H. Tsai, D. T. Lee, and S. Y. Kuo.
4. "Securing Web Application Code by Static Analysis and Runtime Protection", In Proceedings of the 12thInternational World Wide Web Conference (WWW 04), May 2004.
5. Xiang Fu, Xin Lu, Boris Peltsverger, Shijun Chen, "A Static Analysis Framework For Detecting SQL Injection Vulnerabilities", IEEE Transaction of computer software and application conference, 2007.
6. G.T. Buehrer, B.W.Weide and P.A..G.Sivilotti, "Using Parse tree validation to prevent SQL Injection attacks",In proc. Of the 5th International Workshop on Software Engineering and Middleware(SEM '056), Pages 106-113, Sep. 2005.
7. V.B. Livshits and M.S. Lam, "Finding Security vulnerability in java applications with static analysis", In proceedings of the 14th Usenix Security Symposium, Aug 2005.
8. William G.J. Halfond, Alessandro Orso,Panagiotis Manolios, "WASP:Protecting Web Applications Using Positive Tainting and Syntax-Aware Evaluation", IEEE Transaction of Software Engineering Vol34 Nol, January/February 2008.
9. W.G. J. Halfond and A. Orso, "Combining Static Analysis and Run time monitoring to

counter SQL Injection attacks", 3rd International workshop on Dynamic Analysis, St. Louis, Missouri, 2005, pp.1.

10. Marco Cova, Davide Balzarotti, Viktoria Felmetsger, and Giovanni vigna, " Swaddler: An approach for the anamoly based character distribution models in the detection of SQL Injection attacks", Recent Advances in Intrusion Detection System, Pages 63-86, Springerlink, 2007.
11. NTAGW ABIRA Lambert and KANG Song Lin ," Use of Query Tokenization to detect and prevent SQL Injection Attacks", IEEE,2010.
12. Vipin Das 1, Vijaya Pathak2, Sattvik Sharma3 , Sreevathsan4 , MVVNS.Srikanth5,Gireesh Kumar T," NETWORK INTRUSION DETECTION SYSTEM BASED ON MACHINE LEARNING ALGORITHMS", International Journal of Computer Science & Information Technology (IJCSIT), Vol 2, No 6, December 2010.
13. Shui Yu, Member, IEEE, Wanlei Zhou, Senior Member, IEEE,Weijia Jia, Senior Member, IEEE, Song Guo, Senior Member, IEEE,Yong Xiang, and Feilong Tang,(2012), "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient", IEEE Transactions On Parallel And Distributed Systems.
14. Sujatha Sivabalan, Dr P J Radcliffe (2013). "A Novel Framework to detect and block DDoS attack at the Application layer"IEEE 2013-Tencon."
15. Tao Peng and Christopher Leckie and Kotagiri Ramamohanarao (2006), "Survey of Network-based Defense Mechanisms Countering the DoS and DDoS Problems," ACM Transactions on Computational Logic.
16. XIE Yi and YU Shunzheng, "A Detection Approach of User Behaviors Based on HsMM", ITC19/ Performance Challenges for Efficient Next Generation Networks.
17. Yang Xiang, Member, IEEE, Ke Li, and Wanlei Zhou (2011), "Low-Rate DDoS Attacks Detection and Trace back by Using New Information Metrics", IEEE Transactions on Information Technology.
18. Yi Xie and Shun-Zheng Yu, (2009), "A Large-Scale Hidden Semi-Markov Model for Anomaly Detection on User Browsing Behaviors", IEEE/ACM Transactions on Networking.
19. Yu Chen, Member IEEE, Kai Hwang, Fellow IEEE, and Wei-Shinn Ku, Member, IEEE "Collaborative Detection of DDoS Attacks over Multiple Network Domains" IEEE Transactions on Parallel And Distributed Systems.
20. Romil rawat,Shailendra Kumar Shrivastava," SQL injection attack Detection using SVM",ijca, Volume 42– No.13, March 2012.