International Journal of Computer Architecture and Mobility
(ISSN 2319-9229) Volume 4-Issue 6, June 2016

SVM Based SQL-Injection Vulnerability Analysis for SQL Queries

Priyanka Chauhan
chauhanpinku07@gmail.com
Research Scholar, Department of Computer Science and Engineering,
Oriental University, Indore

***Abstract*— Internet is also the supreme offer of information, information is located in distinct information and may be accessible everyplace you go, client can get association at the side of net by method of request covering out there as interface screen, signifies internet is accessible through the utilization of browser during which client may feed his/her information referring to authentication just in case needed through request. Seeing that on read natural setting involving net association and style development, distinct authentication mechanism, security countersign safety and an implausible range of protection treatments are created to defend the approval via unauthorized entry however still crooks square measure aimed towards distinct ways in which to interrupt the particular protection, it should be through hit and walk ways, through infecting system, through stormy system, however at intervals the particular recommend cardstock a recent strategy are offered to urge SQL-Injection being exposed, just in case offered in user's suggestions, the item assessments quandary personal, finger prints and mapping mix to assist suppose any intrusive activities throughout the method, This recommend strategy is easy to use, since it merely wishes process and mapping paradigm involving quandary and every one too straightforward to alter, just in case new personal is found, rather than positioning any overhead at intervals the present doing work method.**

Keywords— SQL-Injection; SVM; Attack;

I. INTRODUCTION

This Injection is that the attack that happens on application layer and solely wants info procedure and flaws in coming up with method. It principally happens owing to improper conductivity of back-end queries, Injection or poisoning has been classified in numerous sorts like SQL-Injection, Script Injection, Shell Injection, XML Injection, hypertext mark-up language Injection[1] etc. it's not possible to shield system for all times by making safe system style once, It should be updated by regular interval of your time, as technology is obtaining advanced multi ways that are offered to trace user, hacker are well trained and behavior primarily based internet-thief ,it takes advantage of mistakes done by either designer or could also be by users.

So for police work and block the attack a simulation tool had been designed, which is able to find internet attacks and additionally block them, and if in future new signature is found may be simply updated within the system. completely different testing procedure are there to check multiple cases however, what's the system isn't designed for them ex: system is intended to find attack[2] of explicit sort it couldn't be ready to find attack of another sort till it's not updated for it.

Attack is hit and trial ways accomplished by many ways, many parameters[3] affects security atmosphere like open input number of tries user gets succeed access, vulnerable atmosphere owing to insecure style and insecure parameter assignment and declaration.DOS Attack, Phishing attack, and a number of other additional attacks ,which are disbursed by causing continuous malfunction request to website for confusing server and flooding them, owing to that server and system gets busy in finding and conniving their legitimacy and creates passing or process of unauthorized question therefore on perform trespasser activities, And phishing attack is carried by making or dynamic internet URL or login credentials ,it additionally carries URL input ID to cause or amendment their values ID, here the aim of assaulter is to execute question at rear. assaulter additionally creates dummy page[4,5],which appearance specifically like original page however once user feeds information within the page it diverts access credentials towards suspicious page, wherever it saves info like username and watchword and question process info to hack user directory page.

Attack may be framed at application layer by providing unwanted info or input, and alternative attack[6] supported network layer ,targets attack by routing ways ,where it routes packet to untargeted approach and so forges the first information to trespasser. here the trespasser diverts the particular traffic to alternative routes and misguides the system by making suspicious id and time delay because it must add alternative 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, and SVM [8] linearly separates data in two separate hyper planes. 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.

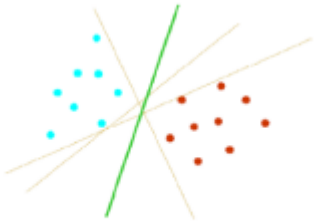


Fig1. Optimal Separating Hyper Plane

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

Livshits and Lam [5] use static analysis techniques to notice vulnerabilities in code. Java Static Tainting uses data flow techniques to notice once tainted input has been wont to create a SQLIA. the first limitation of this approach is that it will notice solely famed patterns of SQLIAs and it will generate a comparatively high quantity of false positives as a result of it uses a conservative analysis.

Java Dynamic Tainting [6] and Securely [7] is another tool that was enforced for java. Despite of alternative tool, chase string rather than character for taint data and check out to sanitize question strings that are generated mistreatment tainted input however sadly injection in numeric fields cannot stop by this approach. problem of distinctive all sources of user input is that the main limitation of this approach.

Two similar approaches by Nguyen-Tong [9] and Pietraszek [8] modify a PHP interpreter to trace precise per-character taint data. A context sensitive analysis is employed to notice and reject queries if bound styles of SQL tokens has been created by illegitimate input. Limitation of those 2 approaches is that they need editing code.

Two approaches, SQL DOM [10] and Safe question Objects [9], use info queries encapsulation for trustable access to databases. They use a type-checked API that cause question building method is systematic. Consequently by API they apply writing best practices like input filtering and strict user input sort checking. The disadvantage of the approaches is that developer ought to learn new programming paradigm or query-development method.

Positive tainting [11] not solely focuses on positive tainting instead of negative tainting however additionally it's machine-driven and would like to developer intervention. what is more this approach edges from syntax-aware analysis, which provides developers a mechanism to manage the usage of string knowledge primarily based not solely on its supply, however additionally on its syntactic role in an exceedingly question string.

IDS [12,19] use AN Intrusion Detection System (IDS) to notice SQLIAs, supported a machine learning technique. The technique builds models of the everyday queries then at runtime, queries that don't match the model would be known as attack. This tool detects attacks with success however it depends on coaching seriously. Else, several false positives and false negatives would be generated.

Another approach during this class is SQL-IDS [13,16] that target writing specifications for the online application that describe the supposed structure of SQL statements that square measure created by the appliance, and in mechanically watching the execution of those SQL statements for violations with regard to these specifications.

A proxy filtering system that intensifies input validation rules on the info flowing to an internet application is termed Security entree [14]. During this technique for transferring parameters from webpage to application server, developers ought to use Security Policy Descriptor Language (SPDL). Therefore developer ought to recognize that knowledge ought to be filtered and additionally what patterns ought to apply to the info.

SQLPrevent [15,20] is consists of protocol request fighter aircraft. The initial knowledge flow is changed once SQLPrevent is deployed into an internet server. The protocol requests square measure saved into the present thread-local storage. Then, SQL fighter aircraft intercepts the SQL statements that square measures created by net application and pass them to the SQLIA detector module. Consequently, protocol request from thread local storage is fetched and examined to see whether or not it contains an SQLIA. The malicious SQL statement would be prevented to be sent to info, if it's suspicious to SQLIA. propose work and Experimental approach 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="abcd"; **Safe Class (O)**

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

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

192.168.1.12 **Safe Class (O)**
192.168.1.15 **Unsafe Class (S)**

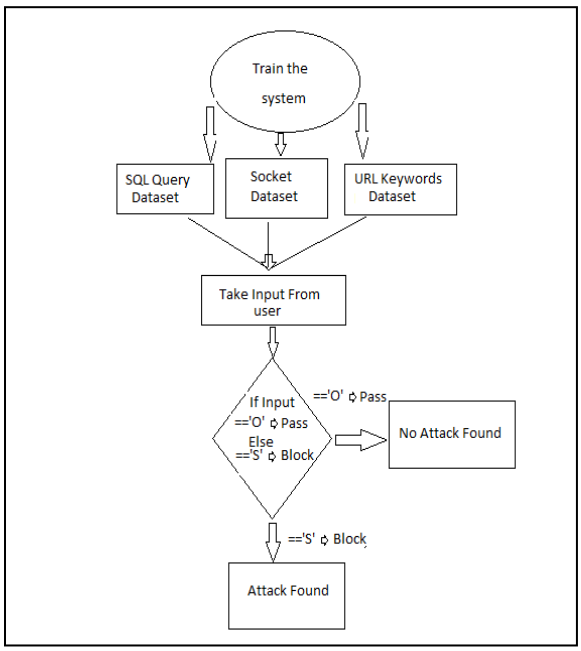
C. Example: ***URL Keywords Dataset***

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

Algorithm:

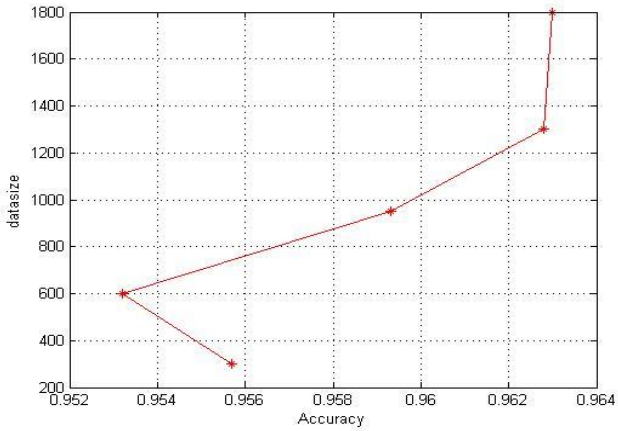
1. *Choose affordable quantity of dataset for training:*
   a) *Select SQL query Dataset*
   b) *Socket Dataset*
   c) *URL Keywords Dataset*
2. *Take input from the user*
3. *Check computer file.*
   a) *If computer file matched with Suspicious category query can blocked*
   b) *If matched with Original category query can pass.*
4. *Calculate totally different parameters (Detection Time, coaching Time, TPR, TNR, FPR, FNR and Accuracy) supported query dismissed.*
5. *method the system for various dataset size, taken to calculate totally different parameter values.*
6. *Repeat the steps one to step five till correct preciseness isn't achieved.*

FLOW CHART

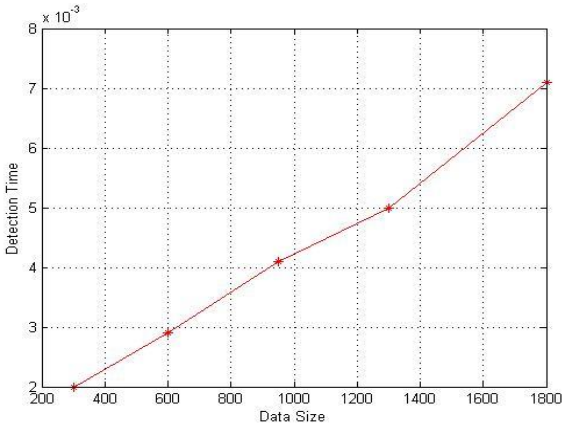


III. RESULT ANALYSIS

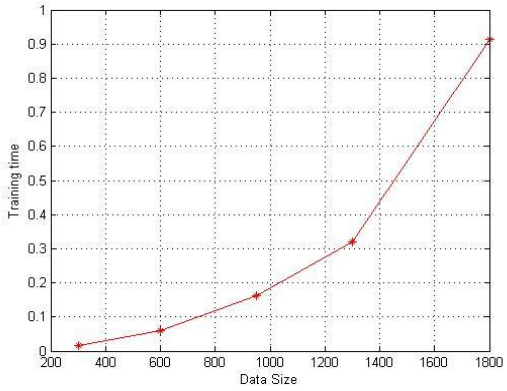
The propose framework has been tested on dataset taken in numerous amount, and totally different parameters has been calculated supported the behavior of dataset size, detection time, coaching time, accuracy. The observation is made on totally different parameters taken a pair of at a time and their fluctuation has been calculated. As dataset of various size has been taken ,it is found that the Accuracy is 96.3% that is best among the offered techniques ,as it may be a lightweight weight system, means that simple to configures ,easy to change ,if new attack signature and finger prints are discovered..



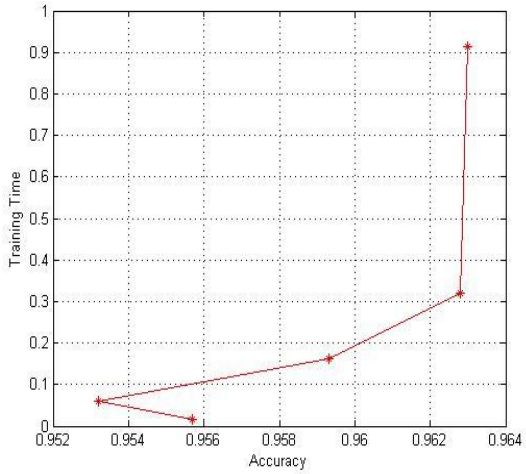
The dataset size of various interval is taken and accuracy is calculated and it's found system is showing near linear behavior.



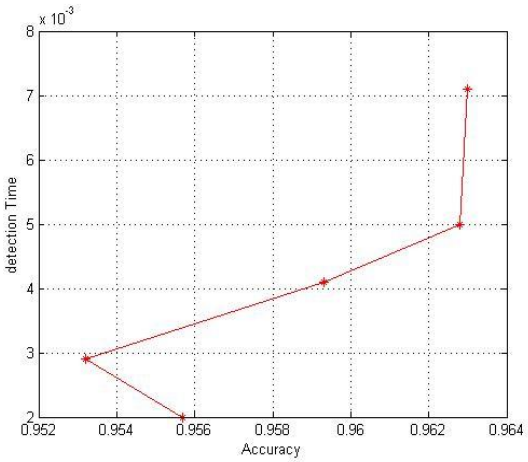
Here, it's clearing predicting that by increasing the dataset size, detection time is additionally increasing, means that the expansion is linear.



The training time increasing once dataset size will increase, because it supported options based mostly system.



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



The training time is ceaselessly variable at regular mounted interval once Detection time varies.

IV. Future Work

The Propose system has been solely outlined for SQL question, Socket question, URL Keywords sorts of attacks, however in future can be increased to trace all multiple attacks by that internet application layer is inclined too. As internet application is directly accessible by legion users, means that legion attacks are potential, that are falling in numerous categories of attacks.

V. Conclusion

The propose system has been conniving completely different marked internet application attacks and categoryifying their sorts supported class provided to that, the system is convenient and lightweight weight and simple to implement on existing system, Sit is taking very little overhead and additionally versatile to feature new signature of dataset ,if found on the system, this system is simply checking SQL question, Socket question, URL Keywords, The system has been tested on completely different dataset size and it's provided accuracy of over 96.3% ,which is found to be best among the all on the market system.

References

[1] William GJ.Halfond, Alessandro Orso," Using Positive Tainting and Syntax Aware Evaluation to Counter SQL Injection Attacks". 14thACM SIGSOFT international symposium on Foundations of software engineering 2006, ACM.

[2] F. Valeur, D. Mutz, and G. Vigna,"A Learning-Based Approach to the Detection of SQL Attacks". In Proceedings of the Conference onDetection of Intrusions and Malware and Vulnerability Assessment (DIMVA), Vienna, Austria, July 2005.

[3] Konstantinos Kemalis and Theodoros Tzouramanis," SQL-IDS: A Specification-based Approach for SQL Injection DetectionSymposium on Applied Computing". 2008, Pages: 2153-2158, Fortaleza, Ceara, Brazil. New York, NY, USA: ACM.

[4] D. Scott and R. Sharp," Abstracting Application-level Web Security". In Proceedings of the 11th International Conference on the WorldWide Web (WWW 2010).

[5] P.Grazie., PhD SQLPrevent thesis. University of British Columbia (UBC) Vancouver, Canada.2008.

[6] Marco Cova, Davide Balzarott," Swaddler: An Approach for the Anomaly-based Detection of State Violations in Web Applications".In Proceedings of the 10th International Symposium on RecentAdvances in Intrusion Detection (RAID), (Queensland, Australia), September 5-7, 2007, pp. 63-86.

[7] S. W. Boyd and A. D. Keromytis," SQLr and: Preventing SQL Injection Attacks". In Proceedings of the 2nd Applied Cryptography and Network Security (ACNS) Conference, pages 292-302. June2004.

[8] W. G. Halfond and A. Orso," Combining static analysis and runtime monitoring to counter Malicious web attacks". In Online Proceeding of the Third International ICSE Workshop on Dynamic Analysis, May 2005.

[9] S. W. Boyd and A. D. Keromytis," SQLRand: Preventing SQL injection attacks". In Proceedings of the 2nd Applied Cryptography and Network Security (ACNS) Conference, Springer-Verlag, June 2004.

[10] C. Gould, Z. Su, and P. Devanbu," Static checking of dynamically generated queries in database applications". In Proceedings of the

26th International Conference on Software Engineering (ICSE'0m,. IEEE Press, May 2004.

[11] Zhendong Su, and Gary Wassermann," The essence of command injection attacks in web applications". In Annual Symposium on Principles of Programming Languages Conference record of the 33rd ACM SIGPLANSIGACT symposium on Principles of programminglanguages, Charleston, South Carolina, USA, 2006.

[12] Ryohei Komiya , Prof. Incheon Paik , "Classification of Malicious Code by Machine Learning".s1140078, University of Aizu, Graduation Thesis. March, 2010.

[13] Amit Kumar Pande," Securing Web Applications From Application-Level Attack", thesis kent university, August 2007.

[14] Peter Scherer, Martin Vicher, Jan Martinovic,"using svm and clustering algorithms in ids system", pp. 108-119, ISBN 978-80-248-2391-1, 2011.

[15] Ankit Anchlia, Sheela Jain," A novel Injection Aware Approach for the Testing of Database Applications", IEEE 2010.

[16] Stephen W. Boyd, Gaurav S. Kc, Michael E. Locasto, Angelos D.Keromytis,and Vassilis Prevelakis," On the General Applicability of Instruction-Set Randomization". IEEE transactions on dependable and secure computing, vol. 7, no. 3, july-september 2010.

[17] Elisa Bertino,Ashish Kamra and James P. Early,"Profiling database applications to detect SQL Injection attacks", IEEE 2007.

[18] R. Ezumalai, G. Aghila," Combinatorial Approach for Preventing SQL Injection Attacks",IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009

[19] Qiujian Zhang, Xiaomei Wang," SQL Injections through Back-end of RFID System", Education Commision of Shanghai, China-06DZ008.

[20] Stephen Thomas and Laurie Williams," Using Automated Fix Generation to Secure SQL Statements", 3rd international workshop on SESS, IEEE 2007..