Automated Vulnerability Detection and Prediction by Security Testing for Cloud SAAS

S. Krishnaveni1*, S. Prabakaran2 and S. Sivamohan3

1Department of Software Engineering, SRM University, Chennai – 603203, Tamil Nadu, India; vanimithila@gmail.com
2Department of Computer Science and Engineering, SRM University, Chennai – 603203, Tamil Nadu, India; prabakaran.mani@gmail.com
3Department of Information Technology, SRM University, Chennai – 603203, Tamil Nadu, India; sivamohan7@gmail.com

Abstract

Background: Cloud SaaS becomes a susceptible target because its shares the application access and data among various tenants. So the careful security testing is necessary to avoid the security problems. As per the recent survey done by OWASP reveals that SQL injection and Cross Site Scripting (XSS) are two of the most serious vulnerabilities in cloud based applications today, because of most dangerous attacks gets exploited and steal the user’s credentials such as cookie, credit card number etc. In order to mitigate these security vulnerabilities cloud service providers adopt a number of vulnerability prediction and detection approaches based on static and dynamic analysis techniques. These existing security mechanisms cannot alleviate attacks targeting cloud based applications which focused on theoretical as well as practical solutions for certain security problems. The developers are necessary to ensure the delivery of safe applications also need to identify potential security issues within the applications before the stored in cloud environment. Method: Security testing is an effective mechanism to find out how vulnerable a Cloud based SaaS may be and to determine whether SaaS offering is susceptible to XSS and resources are protected from intruders. The proposed work provides the automated Security testing by providing vulnerability detection and prediction model. The proposed Hybrid attributes are static and dynamic, the use of dynamic attributes to complement static attributes in the prediction of vulnerabilities. Findings: This paper propose a prediction models by collecting the static and dynamic attributes that are based on classification in order to predict SQLI and XSS vulnerabilities. Improvement: In order to improve the current work we proposed a hybrid attributes such as dynamic attributes to complement static attributes in the prediction of vulnerabilities in cloud based applications. By efficiently using the developed prototype tool for data collection and its used to evaluate our models on six open source cloud based applications. Based on the experiments average results states that the best predictor is MLP which achieved a pd=72%, pf=12% for predicting SQL Injection and pd=82%, pf=19% for predicting XSS Cross Site Scripting. The results of our experiments states that our prediction model is an ease and effective technique to predict and detect the SQLI and XSS vulnerabilities in Cloud applications.

Keywords: Cloud Computing, Data Mining, SaaS Applications, Security Testing, Static and Dynamic Analysis, Vulnerability Detection, Vulnerability Prediction

1. Introduction

As of late distributed computing have seen an eminent development and now it turns out to be most innovation in IT industry. The benefits of utilizing distributed computing are 1. It decreases noteworthy cost in light of the fact that without putting resources into foundation the capacity of figuring can be expanded 2. Its availability, adaptability and versatility. Distributed computing is characterized as “a style of figuring where hugely adaptable IT-empowered capacities are conveyed as a service to outside clients utilizing Internet technologies”. The

*Author for correspondence
different administrations offered by distributed computing engineering are layered as the Infrastructure as a Service (IaaS) giving framework bolster administrations. Platform as a Service layer gives stage arranged administrations, for example, nature for facilitating the client applications. software as a Service (SaaS) layer which have a trademark that it offers a whole application as administration in view of the request of the customer. In SaaS sending model since the information is saved outside the limit of endeavors, so the administration supplier itself give and receive security checks to guaranteeing information security either from assault of programmers and vindictive clients to access unapproved information to them. In the most predominant vulnerabilities in this layer are information infusion vulnerabilities, for example, SQLI and XSS.

In 1,2, Hackers are becomes very smart nowadays, they use new tricky techniques to attack the cloud based web applications to steal the credentials without the user’s knowledge. since cloud computing is open to everyone its avails the resources to many users, so it gets tempting the target to hackers. cloud service providers are giving their software binaries to users, even users may be a hackers, they can easily get in to the cloud and modify the users information. As per the recent Survey done by OWASP reveals that SQL injection and Cross Site Scripting (XSS) are two of the most serious vulnerabilities in cloud based applications today, because of the most dangerous attacks gets exploited and steal the user’s credentials such as cookie, credit card number etc. In order to mitigate these security vulnerabilities cloud service providers adopt a number of vulnerability prediction and detection approaches based on static and dynamic analysis techniques. Security testing is an effective mechanism to find out how vulnerable a Cloud based SaaS may be and to determine whether SaaS offering is Susceptible to XSS and resources are protected from intruders. The proposed work provides the automated Security testing by providing vulnerability detection and prediction model. The proposed Hybrid attributes are static and dynamic, the use of dynamic attributes to complement static attributes in the prediction of vulnerabilities in cloud based applications.

The paper is organized as follows. Section 2 discusses our motivation. Section 3 presents the Proposed Vulnerability Prediction model in this paper. Section 4 presents result and analysis to assess their accuracy based on the proposed hybrid attributes. Section 5 discusses related work. Section 6 presents the summary and concludes our work.

2. Motivation

2.1 Cloud Security Attacks
Cloud computing is still evolving and in continual development, today many software organizations and educational institutions are moving to cloud because they gets the benefits of cloud resources and utilizing distributed computing. Even though peoples enjoy with cloud computing benefits; the security issues in cloud is a key challenges. nowadays hackers are continues to find the tricky techniques to exploits and steal the user’s credentials without the user’s knowledge. Malicious users can inject a malicious code to users websites and gains the unauthorized access to users confidential information stored in cloud. A number of vulnerabilities are remains in cloud computing environment. As per the recent Survey done by OWASP reveals that SQL injection and Cross Site Scripting (XSS) are two of the most serious vulnerabilities in cloud based applications today. these two SQLI and XSS vulnerabilities are occurs due to the improper handling of user’s inputs in cloud based web application programs. SQL injections access the cloud database servers through SQL language, malicious users inject a malicious code in order gain unauthorized access to cloud server databases. In the Figure 1 depicts the SQL injection scenario in cloud environment. Cross-Site Scripting is to defraud the sensitive credentials such as transaction passwords, credit card numbers, etc. of the online user. As shown in Figure 2 represents the Cross-Site Scripting (XSS) scenario in Cloud environment.

Figure 1. SQL injection scenario.
2.2 Vulnerability Trends in Cloud Computing

In order to mitigate the security attacks like SQLI and XSS cloud service providers adopt with various methods and tools, Some of them are keeping away from powerfully produced SQL explanations in code that approve and clean client entered values. As per Symantec MessageLabs Web Security. In Cloud engineering the security of cloud framework lies in two parts a) Multi Layer Security b) URL sifting. Multi Layer Security expects to square malwares and URL sifting lessens means to withdraw from destructive site pages and to ensure data transmission. The study reveals that SQLI is most found shortcoming in SaaS and recommends some bland answers for stay away from SQL Infusion like utilize parameterized questions, approve client contributions by utilizing inbuilt capacities and client characterized work and by utilizing put away methodology. Helplessness Forecast: Powerlessness expectation in view of examining the code or by gathering properties portrays the weakness and building forecast models goes under this zone. In this space specialists relate programming qualities, for example, static code properties with imperfections. The generally utilized static code characteristics are Halstead unpredictability measures and cyclomatic, many-sided quality properties, Lines of Code and so forth. In Figure 3 shows that the number of XSS and SQL injection by years from 2012 to 2016, As predicted, the generalize of XSS and SQL injection vulnerabilities is also noticeable in the Figure 4, As shown in Figure 4 it depicts evolution of SaaS vulnerabilities over the years by types from 2012 to 2017.

3. Proposed Vulnerability Prediction Model

The Proposed Vulnerability Prediction model provide the hybrid attributes such as static and dynamic approaches that based on classifications. according to the research gap identified in the existing approaches are very challenging task in order to amalgamate with cloud based saas applications. Moreover ,we ascertained in existing approaches for prediction of XSS and SQLI Vulnerabilities originated a High rate of false positive and false negatives. In order to improve our Proposed Vulnerability Prediction model we proposed a hybrid attributes such as dynamic attributes to complement static attributes in the prediction of vulnerabilities in cloud based applications. By efficiently using the developed prototype tool for data collection and its used to evaluate our models on six open source cloud based applications. this proposed prediction model has categorized in to two aspects. one is for static and another one is dynamic approaches for collecting the attributes in the web application code. In the static code attributes can be collected by using the program analyzer open source tool called pixy, The dynamic code attributes can be collected by using pattern mining approach, these two
Automated Vulnerability Detection and Prediction by Security Testing for Cloud SAAS

3.1 Hybrid Attributes
The proposed Hybrid approaches based on the static and dynamic analyzers the attribute vector of each sinks in cloud based web application files. A sink denote the node in a control flow graph that may create SQLI or XSS attacks, its consist of two sink SQL sink for database and HTML sink for web client which represents a program statement of a target cloud based web programs. As shown in Table 1 represents the proposed hybrid attributes such as static and dynamic attributes.

3.2 Data Pre-Processing
The Data pre-processing to assure that our proposed prediction model is invariable among datasets with different dispensations. our proposed predictor must knob the different dispensations to hypothesize the results. we intended to used 22 hybrid attributes boycott the target attribute. In this 22 attributes 20 attributes take on numeric values and two attributes are binary. we take the advantage of min–max data normalization techniques to systematize our data. All the training and testing numeric attributes to be normalized. Our prediction model used for zero to one range for prediction.

3.3 Data Reduction
To improve the predictive act of our proposed model we used attribute ranking technique to analyze that attributes that are utmost significant and prove the excepting the attributes are less significant. this data reduction process is called attribute ranking or feature selection techniques. In As per the survey states that there are many techniques have been used by the researcher for attribute ranking or feature selection process. These techniques are such as chi-square, pick up proportion, data pick up, gain ratio etc, we used gain ratio discover the best quality attributes to improve the classifier’s performance.

3.4 Classifiers
The proposed vulnerability prediction model builds on machine learning based classifiers, which classifiers learnt from such data be accurate at vulnerability predictions. These classifiers designate a sensitive sink to a target code that is either vulnerable or non-vulnerable. The data mining tool WEKA is used to implement the classifiers. There are many number of classification algorithms available. In this experiment we used C 4.5 and Multi-Layer Perceptron (MLP) classification algorithms.

3.5 Model Training and Testing
In this Proposed work the dataset is divided into 10 fold cross validation is known as standard sampling method its used for training and testing the two classifiers. The classifier is trained on the 9 folds and then tested on the resting fold, this process is repeated 10 times, to evaluate the classifier effectiveness to predict the vulnerabilities.

3.6 Performance Measures
In order to figure out the best vulnerability predictors, we computed recall or probability of detection (pd), probability of false alarm (pf), and precision (pr). three indicators that are used to determine the performance based on the prediction influence.

- **Accuracy** is measure the sum of True Positives (TP) and True Negatives (TN) over the total number of classifications.
- **Precision** is measure the ratio $\text{PR} = \frac{TP}{TP + FP}$, where FP is the number of false positive it measures the real vulnerable classes that are accurately predicted in terms of a percentage of total number of sinks predicted as vulnerable.
- **Recall** It is measure the ratio $\text{PD} = \frac{TP}{TP + FN}$, where FN is the number of false negatives i.e., measures how good our prediction model is in finding actual vulnerable sink.

4. Result and Analysis
In order to figure out our proposed approach used to establish the investigation with open source cloud based applications with XSS and SQLI vulnerabilities. There are two vulnerability prediction models for predicting SQLI vulnerabilities and XSS vulnerabilities using the collected data. we estimate the two different types of classifiers on the data sets with Vulnerable or Not-Vulnerable. For our experiments the dataset are used from six cloud based applications, they are tadalist, dphoto, flickr, yapig, me, Schoolmate, In Table 2 represents the Datasets collected from cloud based applications, and Table 3 depicts the
accomplishment of SQLI vulnerability predictors and accomplishment of XSS predictors, the overall results were averaged. The data mining tool WEKA is used to implement the classifiers. There are many number of classification algorithms available. In this experiment we used C 4.5 and Multi-Layer Perceptron (MLP) classification algorithms. Based on the experiments average results states that the best predictor is MLP which achieved a pd=72%, pf=12% for predicting SQLI injection and pd=82%, pf=19% for predicting XSS Cross Site Scripting. The results of our experiments states that our prediction model is an ease and effective technique to predict and detect the SQLI and XSS vulnerabilities in Cloud applications.

5. Related Work

This section prospects our work is related to vulnerability prediction, and vulnerability detection approaches. The vulnerability prediction and detection approaches use the machine learning techniques to finding SQLI and XSS vulnerabilities in the target web applications. In [16]
Shar et al. predicted vulnerabilities in open source PHP web applications. They proposed static code attributes approaches to predict the SQLI and XSS vulnerabilities in PHP based web applications, they constructs the prediction model based on machine learning algorithms such as C4.5, NB and MLP, they achieved Recall 86% and pf 3%. In Walden et al. used PHP tokens and software metrics (i.e. LOC, cyclomatic complexity) to predict vulnerability in PHP-MyAdmin, Moodle and Drupal CMS applications. Also they distinguished the software metrics and text mining features and ascertained that text-mining features provide considerably excelling accomplishment in prediction of XSS vulnerabilities. they constructs the prediction model based on Random Forest machine learning algorithm, they attained Recall 80.5% and Accuracy: 75.4% . In Shin et al. used software metrics such as code complexity, code churn, and developer activity metrics to predict vulnerability in Mozilla Firefox Web Browser, Red Hat Enterprise Linux kernel constructs the prediction model based on machine learning algorithms such as Logistic regression, J48, Random forest, NB, Bayesian network and they achieved Recall: 80%. Likewise in Chowdhury et al. used complexity, cohesion and coupling metrics to predict vulnerability-prone files in Mozilla Firefox. They assemble the prediction model based on machine learning algorithms such as Logistic regression, C4.5, Random forest, NB they accomplished

### Table 2. Cloud apps- data sets

<table>
<thead>
<tr>
<th>Test Data Set</th>
<th>#HTMLSink</th>
<th>%Vul XSS</th>
<th>#DBSink</th>
<th>%Vul SQLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>tadalist</td>
<td>120</td>
<td>112</td>
<td>145</td>
<td>132</td>
</tr>
<tr>
<td>dphoto</td>
<td>64</td>
<td>34</td>
<td>72</td>
<td>56</td>
</tr>
<tr>
<td>flickr</td>
<td>133</td>
<td>28</td>
<td>127</td>
<td>67</td>
</tr>
<tr>
<td>yapig</td>
<td>20</td>
<td>40</td>
<td>32</td>
<td>14</td>
</tr>
<tr>
<td>me</td>
<td>73</td>
<td>22</td>
<td>63</td>
<td>35</td>
</tr>
<tr>
<td>Schoolmate</td>
<td>172</td>
<td>138</td>
<td>167</td>
<td>152</td>
</tr>
</tbody>
</table>

### Table 3. Prediction accomplishment of XSS and SQLI

<table>
<thead>
<tr>
<th>DataSet</th>
<th>Classifier</th>
<th>pd</th>
<th>pf</th>
<th>pr</th>
<th>Acc</th>
<th>pd</th>
<th>pf</th>
<th>pr</th>
<th>Acc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>XSS Prediction Result</td>
<td>SQLI Prediction Result</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tadalist</td>
<td>C4.5</td>
<td>48</td>
<td>9</td>
<td>34</td>
<td>75</td>
<td>70</td>
<td>20</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>82</td>
<td>85</td>
<td>76</td>
<td>90</td>
<td>72</td>
<td>12</td>
<td>72</td>
<td>92</td>
</tr>
<tr>
<td>dphoto</td>
<td>C4.5</td>
<td>94</td>
<td>5</td>
<td>84</td>
<td>95</td>
<td>57</td>
<td>8</td>
<td>52</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>85</td>
<td>5</td>
<td>84</td>
<td>93</td>
<td>50</td>
<td>12</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td>flickr</td>
<td>C4.5</td>
<td>60</td>
<td>1</td>
<td>60</td>
<td>60</td>
<td>75</td>
<td>20</td>
<td>75</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>70</td>
<td>20</td>
<td>58</td>
<td>60</td>
<td>100</td>
<td>20</td>
<td>80</td>
<td>89</td>
</tr>
<tr>
<td>yapig</td>
<td>C4.5</td>
<td>97</td>
<td>1</td>
<td>97</td>
<td>99</td>
<td>50</td>
<td>12</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>98</td>
<td>3</td>
<td>95</td>
<td>98</td>
<td>60</td>
<td>8</td>
<td>50</td>
<td>97</td>
</tr>
<tr>
<td>me</td>
<td>C4.5</td>
<td>78</td>
<td>32</td>
<td>80</td>
<td>94</td>
<td>50</td>
<td>12</td>
<td>75</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>93</td>
<td>20</td>
<td>50</td>
<td>82</td>
<td>57</td>
<td>8</td>
<td>50</td>
<td>88</td>
</tr>
<tr>
<td>Schoolmate</td>
<td>C4.5</td>
<td>76</td>
<td>5</td>
<td>99</td>
<td>79</td>
<td>58</td>
<td>15</td>
<td>75</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>99</td>
<td>2</td>
<td>100</td>
<td>99</td>
<td>57</td>
<td>8</td>
<td>50</td>
<td>88</td>
</tr>
<tr>
<td>Average</td>
<td>C4.5</td>
<td>78</td>
<td>23</td>
<td>79</td>
<td>84</td>
<td>58</td>
<td>15</td>
<td>75</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>MLP</td>
<td>82</td>
<td>19</td>
<td>72</td>
<td>79</td>
<td>72</td>
<td>72</td>
<td>12</td>
<td>91</td>
</tr>
</tbody>
</table>
Precision: 4% Recall: 74% Accuracy: 73%, F1 measure: 73%. Scandariato et al. proposed text mining approach machine-learning models for predicting vulnerable files in the K9 mail client application. They used SVM machine learning algorithms to build the prediction model for vulnerabilities in open source Java based web applications. They achieved Accuracy: 87%, Precision: 85%, Recall: 88%. The dominant contradiction between the previous inadequacy prediction methodologies and our proposed work are 1. We contemplate on SQLI and XSS vulnerabilities instead of general vulnerabilities; 2. Subsequently, we propose and make use of static attributes and dynamic attributes opposed to SQLI and XSS; and 3. In this experiment we used C4.5 and Multi-Layer Perceptron (MLP) classification algorithms. Based on the experiments average results states that the best predictor is MLP which achieved a pd = 72%, pf = 12% for predicting SQL Injection and pd = 82%, pf = 19% for predicting XSS Cross Site Scripting. The results of our experiments states that our prediction model is an easy and effective technique to predict and detect the SQLI and XSS vulnerabilities in Cloud applications.

6. Summary and Conclusion

This paper Proposed an automated vulnerability detection and prediction by security testing for cloud saas web applications. This Vulnerability Prediction model provide the hybrid attributes such as static and dynamic approaches that based on classifications. According to the research gap identified in the existing approaches are very challenging task in order to amalgam with cloud based saas applications. Moreover, we ascertained in existing approaches for prediction of XSS and SQLI Vulnerabilities originated a high rate of false positive and false negatives. In order to improve our Proposed Vulnerability Prediction model we proposed a hybrid attributes such as dynamic attributes to complement static attributes in the prediction of vulnerabilities in cloud based applications. Based on the experiments average results states that the best predictor is MLP which achieved a pd = 72%, pf = 12% for predicting SQL Injection and pd = 82%, pf = 19% for predicting XSS Cross Site Scripting. The results of our experiments states that our prediction model is an easy and effective technique to predict and detect the SQLI and XSS vulnerabilities in Cloud applications.

7. References


