INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

AN IMPLEMENTATION OF BROWSER EXTENTION FOR PHISHING DETECTION AND PREVENTION

Archit Shukla*, Lalit Ghelod
* Institute of Engineering & Techology ,M.E. in CS, DAVV, Indore, India

ABSTRACT

The first and greatest casualty of fraud is faith. According to [7] just over two-thirds (68%) of fraud victims say they are less willing to have faith on others after their fraud experience and 63% are less willing to make future investments. As far as, people with criminal intentions are concerned, identity theft is a conventional idea. A con man in police uniform, not cause many victims to become suspicious and they will comply with whatever they are told

Keywords: Aboutsix keywords or phrases in alphabetical order, separated by commas.

INTRODUCTION

Phishing is a model of social engineering techniques used to deceive users phishing is attempting to get information (and sometimes, indirectly, money) such as userid, passwords, by impersonating as a trustworthy entity in an electronic communication. The issue and challenges in the phishing are in the typical use of ssl today, only the server is authenticated, by obtaining an ssl server certificates that is signed by a trusted ca. Ssl also supports mutual authentication, where both the client and the server are authenticated, however this mode of operation requires the user to obtain a personal certificate. A further challenge is how to handle the revocation of credentials. Ssl is designed to prevent eavesdropping, tampering, and message forgery in client/server communications. Instead of attacking the protocol, most phishing attacks use very simple spoofing techniques to trick users into believing that their connection is “secure”. Some phishing attacks exploit the fact that users can not reliably parse domain names (e.g. They can not distinguish ww.paypal.com from www.paypai.com or www.paypal-memberssecurity.com). Many users can not distinguish a legitimate indicator of a secured webpage (e.g. An ssl closed lock icon in the status bar of the browser) from an image of that indicator within the content of a webpage. In many browsers, there is no indicator for “unsecured” sites. so in this paper we are going to implement the web browser extension by which we can easly detect and prevent the phishing attacks.

***Corresponding Author**
Email: Shukla9190@gmail.com

BACKGROUND

Our detection and prevention stratagies were based on the concept of white list, black list , Dns based , url identification and page analysis based. This can be describe in details as . In computing, a **blacklist** or **block list** is a basic access control mechanism that allows through all elements (email addresses, users, URLs, etc.), except those explicitly mentioned. Those items on the list are denied access. The opposite is a white list which means only items on the list are let through whatever gate is being used. A grey list contains items that are temporarily blocked (or temporarily allowed) until an additional step is performed. For example, a company might prevent a list of software from running on its network or a school might prevent a list of web sites from being accesses on its computers. DNS-Based Phishing ("Pharming") Pharming is the term given to hosts file modification or Domain Name System (DNS) based phishing. With a pharming scheme, intruders tamper with a company's host file or domain name system so that requests for URLs or name service return a bogus address and subsequent communications are directed to a fake site.In the url identification based techniques complete url is analysed through decision tree algorithm which match the nodes data with its data base provided in it and provide the result of percentage matching of the url based on that a perticular result can be taken.In the page analysis based detection techniques the redirected page url is analysed and if the page is autenticate one we give indication to them through the indicator signals.

PROPOSED WORK

Cross-Site Scripting (also known as XSS) Is one of the most common application-layer web attacks. XSS vulnerabilities target scripts embedded in a page which are executed on the client-side (in the user's web browser) rather than on the server-side. XSS in itself is a threat which is brought about by the internet security weaknesses of client-side scripting languages such as HTML and JavaScript. The concept of XSS is to manipulate client-side scripts of a web application to execute in the manner desired by the malicious user. Such a manipulation can embed a script in a page which can be executed every time the page is loaded, or whenever an associated event is performed.

XSS is the most common security vulnerability in software today. This should not be the case as XSS is easy to find and easy to fix. XSS vulnerabilities can have consequences such as tampering and sensitive data theft.

KEY CONCEPTS OF XSS

XSS is a Web-based attack performed on vulnerable Web applications, In XSS attacks, the victim is the user and not the application .In XSS attacks, malicious content is delivered to users using JavaScript.

An XSS vulnerability arises when Web applications take data from users and dynamically include it in Web pages without first properly validating the data. XSS vulnerabilities allow an attacker to execute arbitrary commands and display arbitrary content in a victim user's browser. A successful XSS attack leads to an attacker controlling the victim's browser or account on the **Explaining Cross-Site Scripting** vulnerable Web application. Although XSS is enabled by vulnerable pages in a Web application, the victims of an XSS attack are the application's users, not the application itself. The potency of an XSS vulnerability lies in the fact that the malicious code executes in the context of the victim's session, allowing the attacker to bypass normal security restrictions.

IMPACT OF CROSS-SITE SCRIPTING

When attackers succeed in exploiting XSS vulnerabilities, they can gain access to account credentials. They can also spread Web worms or access the user's computer and view the user's browser history or control the browser remotely. After gaining control of the victim's system, attackers can also analyze and use other intranet applications.

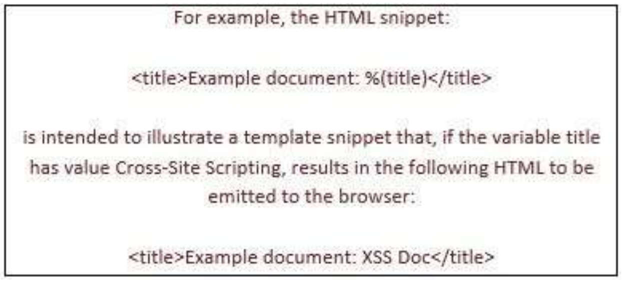
By exploiting XSS vulnerabilities, an attacker can perform malicious actions, such as:

Hijack an account, Spread Web worms, Access browser history and clipboard contents, Control the browser remotely, Scan and exploit intranet appliances and applications, Identifying Cross-Site Scripting Vulnerabilities, XSS vulnerabilities may occur if: Input coming into Web applications is not validated. Output to the browser is not HTML encoded. In recent year's uses of internet are rapidly increases, the number of internet users are also increasing in the same manner. On the other hand internet user now becomes more aware about the internet based frauds and scams. But the numbers of phishing attacks are increases as the internet users are increases and awareness about phishing is increases. To detect and prevent the phishing attacks, the browsers currently usage SSL/TLS, but these techniques is not much effective and still allows web spoofing, i.e. misleading users by impersonation or misrepresentation of identity or of credentials. Indeed, there is an alarming increase in the amount of real-life web-spoofing attacks, usually using simple techniques. Often, the swindlers lure the user to the spoofed web site, e.g. impersonating as financial institution, by sending her spoofed E-mail messages that link into the spoofed web-sites; this is often called a phishing attack. The goal of the attackers is often to obtain user-ID's (Identity), passwords/PIN (Personal Identification Number) and other personal and financial information, and abuse it e.g. for identity theft. Thus, the significant improvement in detection of spoofed sites is required.

PROPOSED SOLUTION

To overcome described problem given above a solution is proposed in this section of document. The proposed approach of a browser extension i.e. plug in. the proposed plugin is a hybrid approach which is derived using various individual techniques. Moreover it solution also includes usability experiments, to measure and compare the effectiveness of the approach to sites identification indicators, the following work provide the given solutions:

1. Provide prevention from the spoofing and phishing done by using a URL.
2. Allow to open, but the authentic web pages.
3. Provides the authenticity rating to the web pages.
4. Implement a novel technique to detect and propagate the cross site scripting using browser extension
5. Maintain the data base according to which authenticity rating will be provided.



For example, the HTML snippet:

<title>Example document: %(title)</title>

is intended to illustrate a template snippet that, if the variable title has value Cross-Site Scripting, results in the following HTML to be emitted to the browser:

<title>Example document: XSS Doc</title>

Table (1) XSS Examples

IMPLEMENTATION

The desired systems architecture is given using figure 1 in this system architecture all the subsystems are performing the intermediate task for providing the validity of web page.

Internet explorer: That is Microsoft web browser; user put their URL as input for navigating the page.

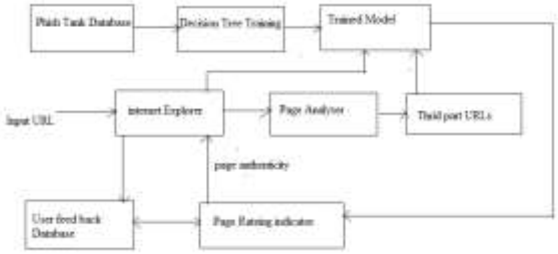


Figure 1 System Architecture

Phish tank database: That is data base contains various phishing URLs that reported by different users and the organizations as a phishing URL.

Decision tree Training: There are two different decision trees are listed and using the phish tank database these models are trained.

Trained model: Here the trained model is stored for detecting the phishing URLs and patterns. Using the input URL this trained model provides their authenticity.

Page analyser: In this phase the web page is analysed and the third part URL is collected from the web pages. These web page URLs are evaluated using trained data model and the user feedback data base for their validity.

User feedback data base: User feedback database is collection of user experience about the behaviour of web content.

Web page rating indicator: Using the third party URL analysis, navigated web page analysis and the past user experience the web page validity is approximated for providing the web page authenticity using their rating bar.

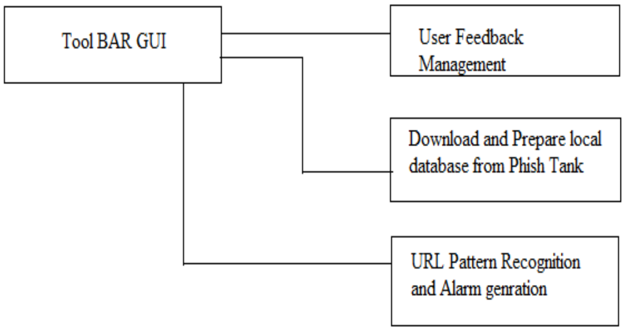

Tool BAR GUI
User Feedback Management
Download and Prepare local database from Phish Tank
URL Pattern Recognition and Alarm genration

Fig 2 Shows The GUI Planning

- **Toolbar GUI:** It is an effort to demonstrate the user interface to show different notifications and there information generated by the system. Actually it is individual software unit to obtain the URL which is typed over the internet explorer and it explores the information regarding the collected URL.
- **User feedback management:** Here user can add the positive or negative feedback for particular URL that is updated over a server. By which different toolbar users can access the user feedback from the server.
- **Prepare local database from phish tank:** Phish tank is a large database of phishing web sites and it contains different URLs which is found as fraud or phishing work.
- **URL pattern recognition:** Were a new concept implemented to determine the URL patterns of phishing using a decision tree algorithm that help us to predict is a URL is a phishing URL or not.

The first and greatest casualty of fraud is faith. According to [7] just over two-thirds (68%) of fraud victims say they are less willing to have faith on others after their fraud experience and 63% are less willing to make future investments. As far as, people with criminal intentions are concerned, identity theft is a conventional idea. A con man in police uniform, not cause many victims to become suspicious and they will comply with whatever they are told

Keywords—Aboutsix keywords or phrases in alphabetical order, separated by commas.

The first and greatest casualty of fraud is faith. According to [7] just over two-thirds (68%) of fraud victims say they are less willing to have faith on others after their fraud experience and 63% are less willing to make future investments. As far as, people with criminal intentions are concerned, identity theft is a conventional idea. A con man in police uniform,

not cause many victims to become suspicious and they will comply with whatever they are told

Keywords—Aboutsix keywords or phrases in alphabetical order, separated by commas.

RESULT ANALYSIS

The given chapter of the document includes the performance of the classifiers that are implemented in the current anti-phishing browser extension. Chapter includes various performance parameters that are required to evaluate for performance analysis, provided results are the classification performance due to continuously increasing data.

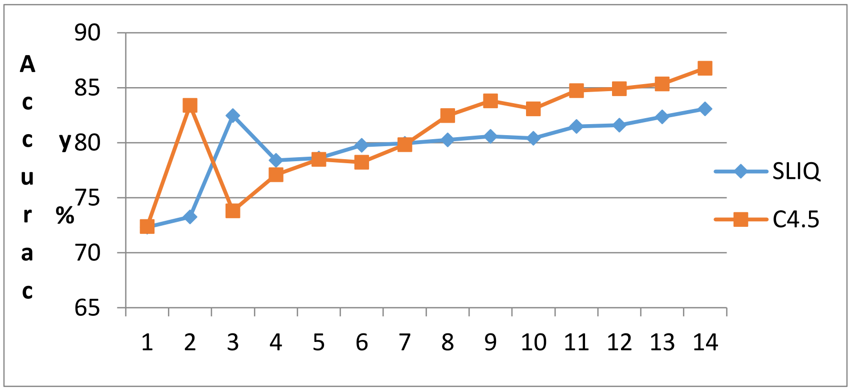


Figure 1 Accuracy

Classification Accuracy

The calculated performance of the proposed system in terms of accuracy is measured in terms of percentage which is evaluated using n cross validation method.The overall classification accuracy is given below and evaluated using the below given formula, the listed accuracy of the system are the best performance during different experiments.

$$\%\ \text{accurecy} = \frac{Total\ correctly\ classified}{total\ values\ to\ classify} X10$$

Error rate

The amount of misclassified instances is given as the error rate of the system. That can be calculated using the below given formula

$$\%\ error = \frac{Total\ incorrectly\ classified}{Total\ Number\ of\ objects} X100$$

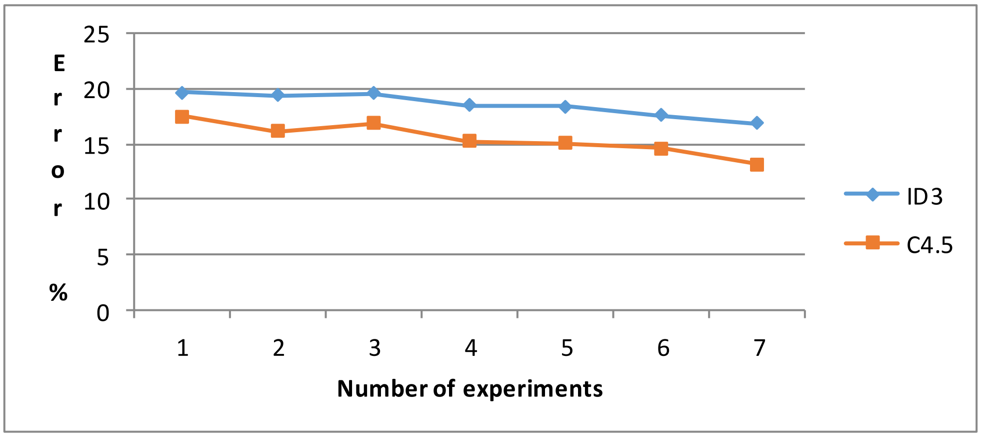


Fig.2 Error rate

During the different experiments the performance of the classifiers are remains fixed, therefore the system performance is stable during variations of data. The error rate of the system is given using figure 5.2, in this diagram the X axis provides different experiments and the Y axis provides the Error rate of the system.

Memory consumption

That is defined as the memory resources consumed during the performance of the system, here the consumption of the resources in giving in KB.

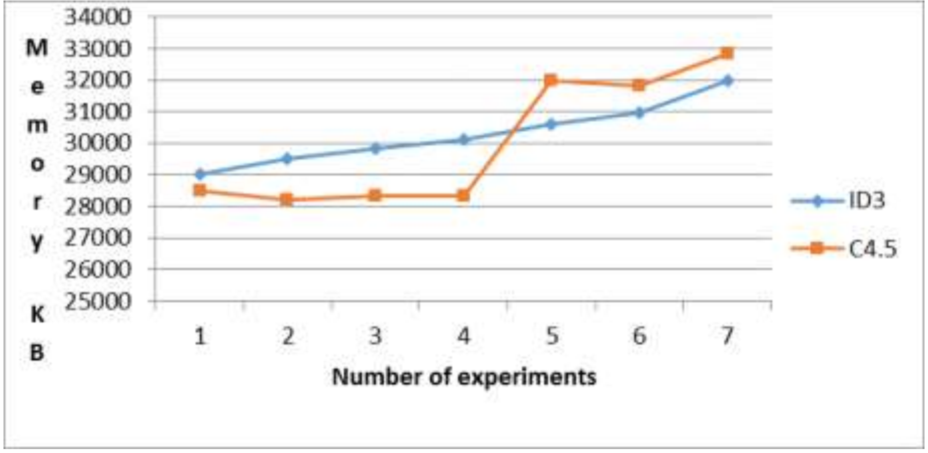


Figure 3 Memory Consumed

Memory consumption of both classifiers are consumes about similar performance and increasing as the size of data in main memory is increases.

Model Building Time

That is the amount of time required to build data model (algorithm training time), which is evaluated in terms of Millisecond

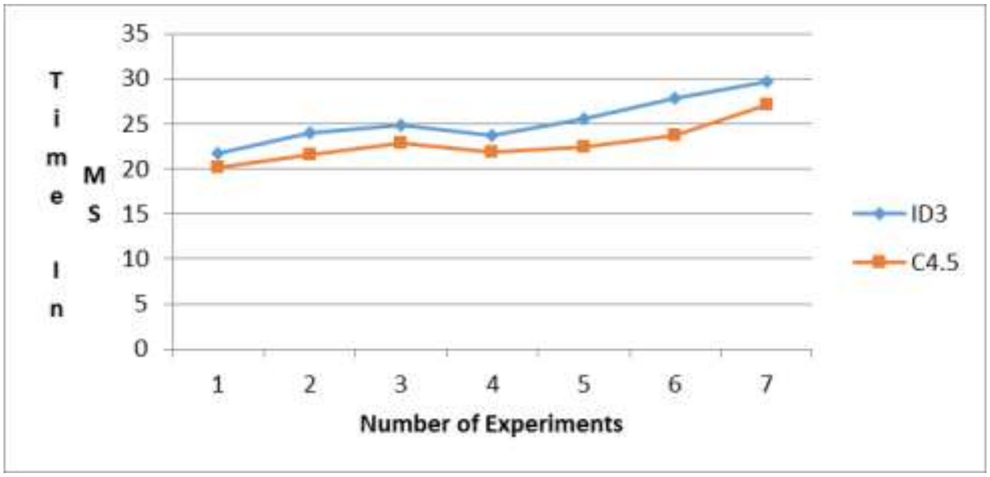
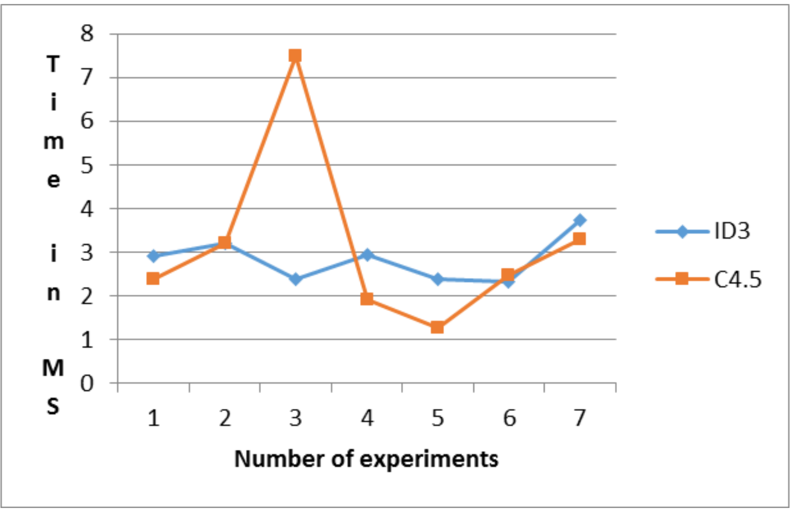


Figure 4 Model Building Time

Decision Time

The amount of time needed to reach a correct decision after training for classified value prediction is given using given figure 5.5 search time of any predictive algorithm is estimated for all the samples which is produced in validation of the trained model. The decision time of any decision tree algorithm is depends upon the depth of generated tree, and the intermediate node



.

Figure 5 Decision Time

CONCLUSION

The proposed study is an investigation about the phishing attack and their social effects. After analysis of different attack techniques and probability that found that, phishing attack is a critical kind of attack. Using this attack a phisher damage the socially and financially .For that purpose various detection and prevention technique are also investigated. By which a new solution is discovered. The proposed system incorporates the machine learning techniques, white and black list concept and user feedback methodology for detecting and avoiding the phishing attacks.

The proposed system is implemented using visual studio dot net framework, and the performance is evaluated using the n cross validation technique. After performance analysis the system provides the effective results that are summarized using figure 6.1.

S. No.	Parameters	Remark
1	Detection accuracy	The machine learning models are able to classify the data more appropriately and accurately the overall performance is adoptable
2	Time complexity	The model build time and search time of the system is low therefore less time complexity of system make it more adoptable
3	Space complexity	The system consumes less memory resource during phishing analysis, therefore the system is adoptable

The presented result of the system is adoptable and efficient, due to their pattern analysis capability, less memory and time consumption during processing and analysis.

ACKNOWLEDGEMENT

The problem of Phishing does not have a single solution as of today. Phishing is not just a technical problem and Phishers would keep coming up with new ways of attacking the users. Online users should undertake periodic vulnerability analysis to identify and plug weaknesses that can lead to a successful Phishing attack. To guard against these threats, user need to be educated on the dangers of advanced malware and the forms it can take today. In addition, security teams need advanced technologies that can detect and stop the advanced threats that are currently bypassing their conventional defenses. The proposed model is efficient and working as desired, that is adoptable due to their accurate URL classification and less computational resource consumption. The provided results demonstrate the effectiveness of system. That is more extendable with global database creation using cloud and with DNS verification methodology is possible.

REFERENCES

1. *"Phishing: Challenges and Issues in Malaysia" Madihah Mohd Saudi, Islamic Science University of Malaysia (USIM), Negeri Sembilan, Malaysia Shaharudin Ismail, Islamic Science University of Malaysia (USIM), Negeri Sembilan, Malaysia Emran Mohd Tamil, University Malaya, Malaysia Mohd Yamani Idna Idris, University Malaya, Malaysia.*
2. *Online Detection and Prevention of Phishing Attacks (Invited Paper)" Juan Chen Institute of Communications Engineering Nanjing 210007, P.R. China icechj@msn.com Chuanxiong Guo Institute of Communications Engineering Nanjing 210007, P.R. China xguo@ieee.org*
3. *"Learning to Detect Phishing Emails" Ian Fette School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA icf@cs.cmu.edu Norman Sadeh School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA Anthony Tomasic School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA*
4. *"A Comparison of Machine Learning Techniques for Phishing Detection" Saeed Abu-Nimeh1, Dario Nappa2, Xinlei Wang2, and Suku Nair1 SMU HACNet Lab Southern Methodist University Dallas, TX 75275*
5. *NISR The Phishing Guide Understanding & Preventing Phishing Attacks*
6. *"Protecting Users Against Phishing Attacks with AntiPhish" Engin Kirda and Christopher Kruegel Technical University of Vienna*
7. *Canadian Securities Administrators | www.csa-acvm.ca ,Innovative Research Group, Inc. | www.innovativeresearch.ca*

8. *Proactively Stop Web Based Fraud and Fraudlent Transaction With ThreatRadar Fraud Prevention, www.imperva.com*
9. *Modeling and Preventing Phishing Attacks by Markus Jakobsson, Phishing detection system for e-banking using fuzzy data mining by Aburrous, M. ; Dept. of Comput., Univ. of Bradford, Bradford, UK ; Hossain, M.A. ; Dahal, K. ; Thabatah, F.*
10. *Learning to Detect Phishing Emails : Authors AnIan Fette (Carnegie Mellon University),Norman Sadeh (Carnegie MellonUniversity),Thony Tomasic (Carnegie Mellon University*