MULTI-LEVEL PARSING BASED APPROACH AGAINST PHISHING ATTACKS WITH THE HELP OF KNOWLEDGE BASES

Gaurav Kumar Tak[1] and Gaurav Ojha[2]

[1]School of Computer Science and Information Technology, Lovely Professional University, Phagwara, Punjab – 144402, India

[2]Department of Information Technology, Indian Institute of Information Technology and Management, Gwalior, Madhya Pradesh - 474010, India

ABSTRACT

The increasing use of internet all over the world, be it in households or in corporate firms, has led to an unprecedented rise in cyber-crimes. Amongst these the major chunk consists of Internet attacks which are the most popular and common attacks are carried over the internet. Generally phishing attacks, SSL attacks and some other hacking attacks are kept into this category. Security against these attacks is the major issue of internet security in today's scenario where internet has very deep penetration. Internet has no doubt made our lives very convenient. It has provided many facilities to us at penny's cost. For instance it has made communication lightning fast and that too at a very cheap cost. But internet can pose added threats for those users who are not well versed in the ways of internet and unaware of the security risks attached with it. Phishing Attacks, Nigerian Scam, Spam attacks, SSL attacks and other hacking attacks are some of the most common and recent attacks to compromise the privacy of the internet users. Many a times if the user isn't careful, then these attacks are able to steal the confidential information of user (or unauthorized access). Generally these attacks are carried out with the help of social networking sites, popular mail server sites, online chatting sites etc. Nowadays, Facebook.com, gmail.com, orkut.com and many other social networking sites are facing these security attack problems.

This paper discusses a Knowledge Base Compound approach which is based on query operations and parsing techniques to counter these internet attacks using the web browser itself. In this approach we propose to analyze the web URLs before visiting the actual site, so as to provide security against web attacks mentioned above. This approach employs various parsing operations and query processing which use many techniques to detect the phishing attacks as well as other web attacks. The aforementioned approach is completely based on operation through the browser and hence only affects the speed of browsing. This approach also includes Crawling operation to detect the URL details to further enhance the precision of detection of a compromised site. Using the proposed methodology, a new browser can easily detects the phishing attacks, SSL attacks, and other hacking attacks. With the use of this browser approach, we can easily achieve 96.94% security against phishing as well as other web based attacks.

KEYWORDS

Phishing, SSL, Artificial Intelligence, Internet Attacks, Knowledge Base, Parsing

1. INTRODUCTION

In present world, Internet plays a very important role in everyone's daily life. Nowadays any work can be performed over the internet. Just to name a few Internet Banking, Online Ticket

Booking, Hotel Booking, Social Networking, Online Shopping etc are getting more popular day by day. Hence security over the internet is of utmost importance in today's scenario.
On the World Wide Web, Cyber-crime is one of the major security issues, troubling internet security. These crimes can be defined as immoral actions performed with the use of internet. They include illegal access of data, illegal interception of data, eavesdropping of authorized data over an information technology infrastructure , data interference (which includes unauthorized damaging, deletion, deterioration, alteration or suppression of computer data), Unethical access of web services, Disturbance of social-peace, systems interference (interfering with the functioning of a computer system by inputting, transferring, destroying, removing, deteriorating, altering or suppressing computer data), misuse of devices, forgery (ID theft), and electronic fraud.[16][13]
Some of the cyber-crime issues have become high-profile, particularly those surrounding hacking, copyright infringement, child pornography and child grooming.

However in the context of internet security, phishing is the most commonly used web attack. Phishing can be defined as the fraudulent process of masquerading as a trustworthy entity in an electronic communication so as to acquire sensitive user information (such as usernames, passwords) and other confidential information (like security key and credit card or debit card details, master card details). In phishing attacks often unsuspecting users are lured using communications purporting to be from popular social web sites, auction sites, online payment Gateway or IT administrators. These attacks are usually carried out by e-mail or instant messaging in which the users are directed to a fake website whose look and feel are almost identical to the legitimate one. Here the user is prompted to enter personal details which go directly into the hands of the cyber criminals. Even while using server authentication, it may require tremendous amount of time and skills to establish that the website is fake. Phishing is an example of social engineering techniques used to fool users by exploiting the poor usability of current web security technologies. It can be used to break the security system of many web services, to access many authorized information unethically.

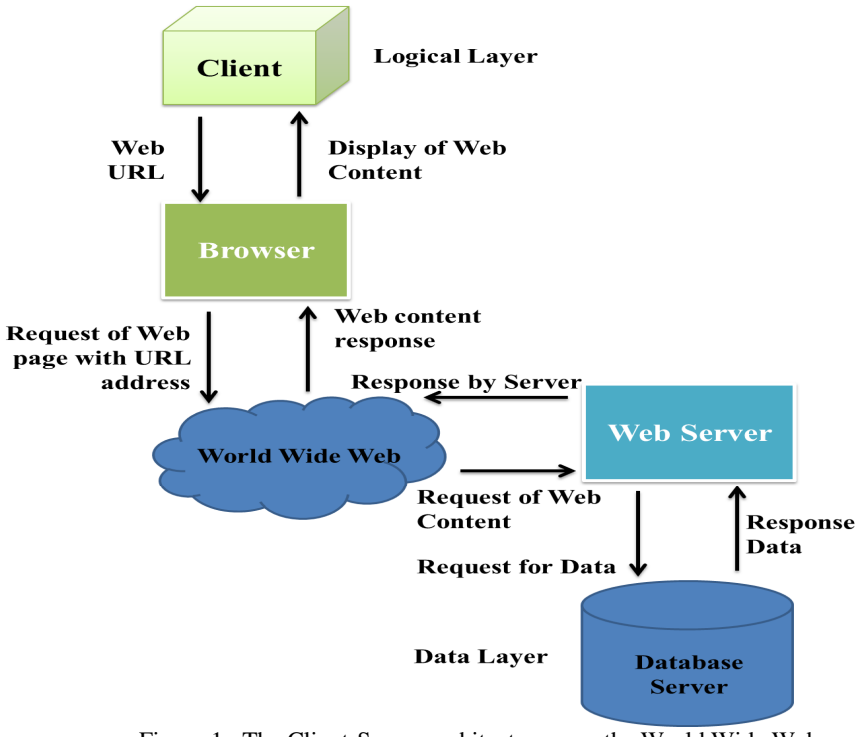


Figure 1. The Client-Server architecture over the World Wide Web

Most of the services over the Internet are mainly based on the client–server architecture which provides communication all over the World Wide Web. The Client–server model of computing is a sort of distributed application structure. It divides the whole of the workload between different service providing stations and service requester stations, known as servers and clients respectively. Generally, when a client requires accessing some web pages, then it reaches to the server with the help of a web browser. Result of the request also follows the same process but in reverse order. A server machine is a host which is running one or more server programs sharing its resources with clients. A client shares none of its resources. However it can always request a server's content or service function. Hence, clients initiate communications with servers when needed.

Phishing attack can be defined as an attempt by a person or a group of people to steal some information (for security purpose) such as user ID, passwords, credit card information, etc. from unsuspecting victims for identity theft, financial gain or other fraudulent activities. Fake websites which looks very similar to the genuine ones are hosted to achieve this. Many a times it is too difficult to differentiate the fake from the genuine one. Thus more often than not the internet users assume that they are entering data into a genuine website without realizing that they are giving away their precious information to a phishing attacker who can misuse it for many purposes according to his convenience. Architecture of phishing attack is shown in Figure 2 below.

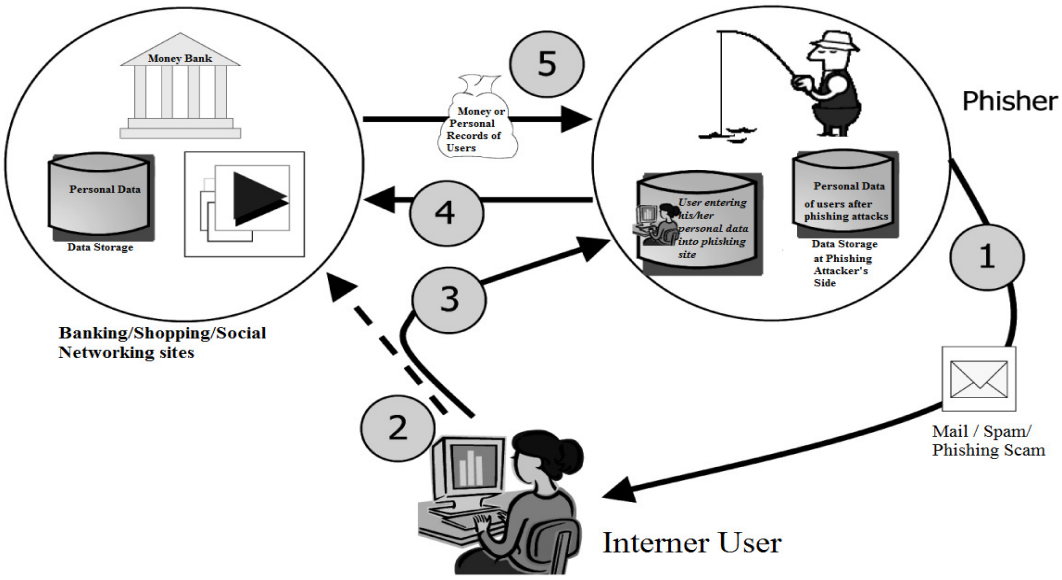


Figure 2. Architecture of Phishing Attacks

In this document, we are proposing a new technique for stopping phishing attacks by introducing the concept of parsing the web-URL (Uniform Resource Locator) before visiting it. The technique also proposes the use of knowledge base to retrieve some information that is stored previously. Using the Knowledge Base, we can gain the better security against phishing attacks and reduce the time complexity of the operations. Multi parsers are used for multiple operations, hence easing detection of the phishing attacks. Here in this methodology the browser will be more participating in the process of detecting and prevention of phishing attacks.

A Knowledge Base is the modelling of previously occurred events in order to predict future events by employing some artificial intelligence techniques [5]. It is a sort of database for knowledge management, providing the means for the computerized collection, organization, and

retrieval of knowledge. Also a collection of data representing related experiences, their results is related to their problems and solutions.
They are basically artificial intelligent tools providing intelligent decisions. Knowledge is obtained and represented using various knowledge representation techniques rules, frames and scripts. The basic advantages offered by such system are documentation of knowledge, intelligent decision support, self-learning, reasoning and explanation. [6]

Each knowledge base follows the DIKW chain processing in its thinking and reasoning process. The Chain consist four elements which are as follows: data, information, knowledge and wisdom. All elements are different in nature and have their own impact at the time of decision making. Data concern with the observation and some raw facts. Data are meaningless without an additional processing viz. filtering, comparing etc. Information can be defined as the processed data. In short, Knowledge is defined as follows: knowledge is an outcome of processes like synthesis, filtration, comparison and analysis of information which are already available with the knowledge base and to produce meaningful and logical results. The elements of the chain can be arranged as shown in Figure 3.

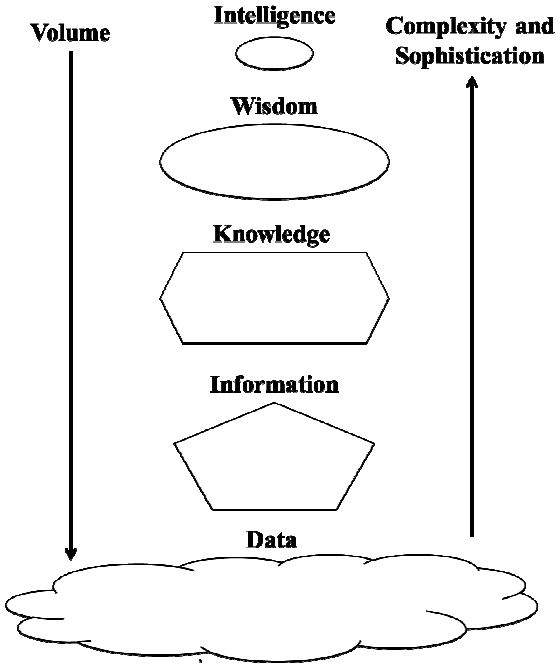


Figure 3. The DIKW Chain

As we all know that phishing attack is a URL based attack which happens between the Internet user and the browser, so our proposed methodology gives the new security layer between browser and the User using the Knowledge Base and some parsing operations.

2. RELATED WORK

Many techniques and algorithms have been developed and implemented for prevention of phishing and to secure the theft of confidential information (usernames, passwords, security key, credit card /debit card/master card details). But there are still many issues remaining on this matter.

Many techniques and schemes are being proposed to provide a secure environment for e-banking services, e-commerce services and payment gateway services and to block the sniffing,

eavesdropping etc. So that transmission of the confidential information will be preserved and unauthorized personnel can't access that information.
But day by day, phishing attacks are increasing. While most phishing attacks target the financial transaction website (Banking site, e-commerce, e-shopping website, payment gateway websites), more and more phishing incidents targeting online game operators and large ISPs (internet service provider) have also been discovered.

Many approaches (e.g. toolbars) have been proposed to prevent phishing attacks. The anti-phishing toolbars is also a common but not so user friendly approach out of them. It is based on web browser plug-ins that warns browsers when they visit any suspected phishing site. Commonly, anti-phishing tools use two major approaches for mitigating phishing sites. The first approach is based on heuristics to check the host name and the URL for common spoofing techniques. The second method lists out some blacklist phishing URLs. The heuristics approach is not 100% accurate since it produces low false negatives (FN), i.e. a phishing site is mistakenly judged as legitimate, which implies they do not correctly identify all phishing sites. The heuristics often produce high false positives (FP), i.e. incorrectly identifying a legitimate site as fraudulent. Blacklists have a high level of accuracy because they are constructed by paid experts who verify a reported URL and add it to the blacklists if it is considered as a phishing website. [16][13][9]
Delayed password disclosure [7] is another new method to avoid phishing attacks. This method discusses a user interface that checks the authenticity of the website as the user enters his/her password. This is based on the feedback generated by the interface as user enters the password; hence if the feedback generated is not according to the authentic website an alarm is triggered.
Another method to create awareness amongst users against phishing is Trust bar construction [8]. This method associates logos with the public key of the website being visited hence easing the way of authentication of website. Passmark™ is a similar method currently being used by Bank of America. This method fights phishing by authenticating the website back to the user. Here the website first identifies user by previous cookies and before the password submission it sends back a user specific image. If the user identifies the image then and only then he should enter the password.

The detection and identification of phishing websites in real-time, particularly for e-banking/payment gateway website, is a very complex and dynamic problem which involves many factors and criteria. Many methods like improving site authenticity, one time password, having separate login and transaction password, personalized e-mail communication, user education about phishing are being implemented to prevent phishing attacks, but they don't provide high security.

3. PROPOSED METHODOLOGY

The proposed browser based methodology against phishing attacks utilizes some of the basic information of the domains. For instance, more often than not a phishing website will be a newly registered domain. Furthermore they will have some identical portion of the legitimate website domain. Here we propose a knowledge base approach against phishing attacks which also uses some parsing techniques to detect the attack.

3.1. Knowledge Bases

Our methodology uses some knowledge bases which are described as follows:

3.1.1. Knowledge Base I

Knowledge Base Initial or KBI stores the pattern and other detection methods of previously detected phishing attacks and other web attacks. It validates the URL and also relates the URL with the previously detected phishing attacks. If pattern of new URL matches with the previously stored Phishing attacks, then it generates a phishing alert before visiting the URL. Since KBI only stores the recent and frequently occurring phishing attacks so the size of Knowledge Base I can vary according to the requirements of the situations and the security threats posed in the scenario. This is also named as Knowledge Base Initial because it is used in the beginning of the methodology.

3.1.2. Knowledge Base T

Knowledge Base Trusted or KBT maintains all the trusted and secure URLs which are previously visited on the same browser. The user can further manually add the frequently visited legitimate websites to this knowledge base for whom he wishes not to carry out security checks every time. If the URL is present in this knowledge base then it is deemed secure. Else if the URL would be considered to lie in the danger zone of security then all the security analysis will take place for that URL before visiting it.

3.1.3. Knowledge Base A

This Knowledge Base defines all the URL-pattern based phishing and SSL attacks which have detected previously by the browser till date. This Knowledge Base is used before the operation of ‘Parser-1’.

3.1.4. Knowledge Base B

This Knowledge stores the all information (like license year, rating of the domain, popularity of the domain etc.) of the URLs which is previously visited and detected as the Phishing attacks.

3.1.5. Knowledge Base C

This Knowledge Base defines all the URL-pattern based phishing and SSL attacks which have detected previously by the browser till date. This Knowledge Base is used before the operation of ‘Parser-1’.

3.1.6. Knowledge Base D

This Knowledge Base stores all the URLs which are previously visited. It is responsible for maintaining the history of the all the URLs which are previously visited. This knowledge base is already a common feature in almost all the web browsers.

Currently most of the browsers (like Mozilla Firefox, Internet Explorer, Opera, etc.) maintain the history of the previously visited URLs. When an internet user types the URL keywords in the address bar of browser then it automatically suggests all the URLs pertaining to that keywords which were frequently visited using this Knowledge Base (history of URLs). Figure 4 below represents all suggested URLs by the browser ‘Mozilla Firefox’.

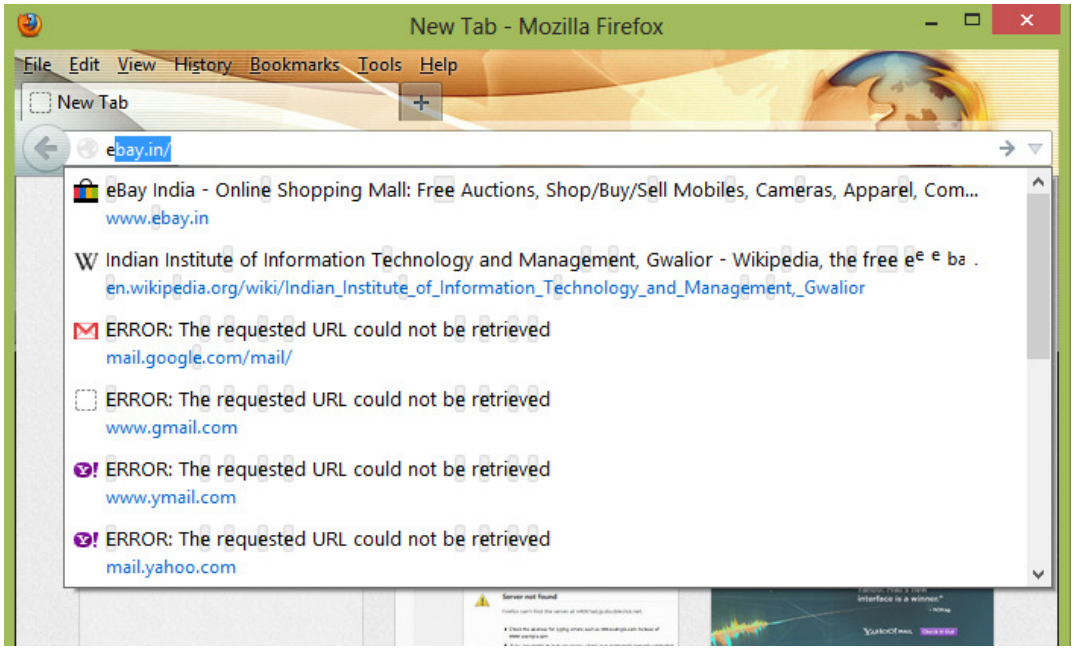


Figure 4. The browser shows a list of previously visited URLs (Firefox 19)

3.2. Parsers

Some Parsers are also used in the detection of URL based attack in the proposed methodology, which are described as follows:

3.2.1. Parser 1

It is used to detect the pattern based URL attacks. This parser provides the security against phishing attacks as well as SSL attacks. It also analyzes the usage of some special character (like '-','.' etc.) in the URL to detect the attacks. This parser's operation is based on the fact, that phishing attackers use the some fraction of the actual legitimate URL so as to generate a close to real phishing URL.

For example take the URL http://www.firstgenericbank.com.account-updateinfo.com/, it is a phishing URL of First Generic Bank. The user can be fooled to believe this is a legitimate website as it contains part of the original URL separated by 4 dots. The following Figure 5 represents the Phishing attack example over the generic bank website.

3.2.2. Parser 2

When a URL is parsed into this, all the details of the website such as license year, rating of the domain, popularity of the domain etc. become available to the browser. Using these details parser-B can declares if the URL is phishing website URL or a legitimate website URL. This parser takes account of the fact that phishing URLs are newly registered one with low rating and popularity. Hence if the URL is newly registered, then it can be a phishing attack on any existing URL.is used to detect the pattern based URL attacks. Some Browser (Like Internet Explorer 7.0, Opera etc.) also use this approach for the detection of web attacks. Internet Explorer 7.0 browser also use the site rating to detect the phishing websites, but many a times it is not user friendly and

unable to detect all attacks. Figure 5 represents the information of URL http://www.facebook.com in Opera Browser.

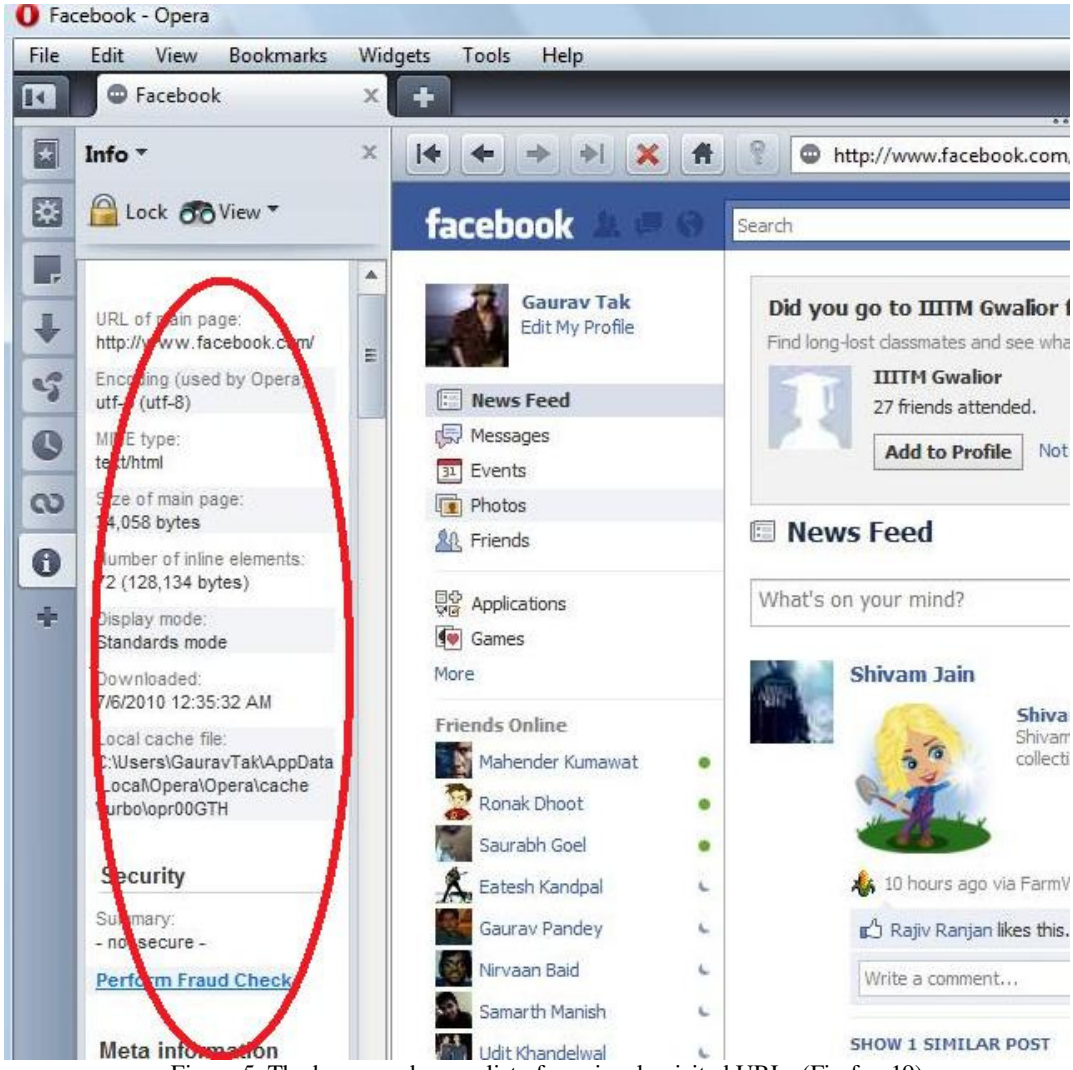


Figure 5. The browser shows a list of previously visited URLs (Firefox 19)

3.2.3. Parser 3

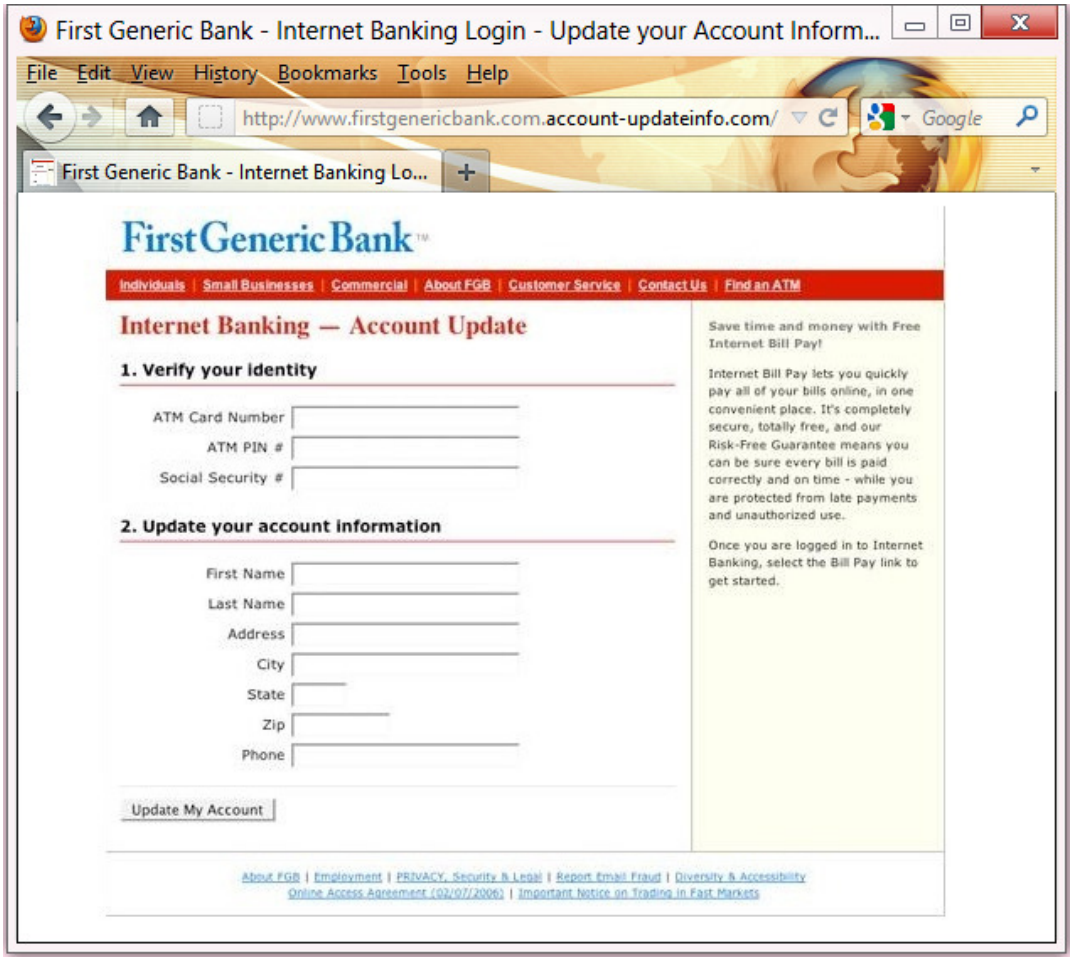
This parser performs an important step for security against the phishing attacks. It performs the fraud check analysis of an URL and generates a warning message if URL is not secure.

3.2.4. Parser 4

It searches for other URLs whose pattern matches with the requested URL. It finds all details of the other similar URLs and compares all the details (like year of domain registration, rating of the domain, popularity of the domain etc.) with the requested URL details. It then displays all the results in the preference on the browser screen before visiting the requested URL.

Parser 2 and Parser 4 act like web crawlers and scan the World Wide Web to get the required information to detect the phishing as well as other web attacks. Both parsers work in automated

manner, they also provide the indexing and relative weights to compare the outcomes. Some



policies have been also defined for both parsers for their crawling operation.

Figure 6. A sample phishing website

In implementation of parser 4 and 5, the Open Source Crawler "crawler4j" has been used. The java code of the "crawler4j" is as follows –

```
import java.util.ArrayList;
import java.util.regex.Pattern;
import edu.uci.ics.crawler4j.crawler.Page;
import edu.uci.ics.crawler4j.crawler.WebCrawler;

import edu.uci.ics.crawler4j.url.WebURL;

public class MyCrawler extends WebCrawler
{
Pattern filters = Pattern.compile(".*(\\.(css|js|bmp|gif|jpe?g"
                + "|png|tiff?|mid|mp2|mp3|mp4"
                + "|wav|avi|mov|mpeg|ram|m4v|pdf"
                + "|rm|smil|wmv|swf|wma|zip|rar|gz))$");
}

public MyCrawler()
```

```
{
      }

public boolean shouldVisit(WebURLurl)
{
     String href= url.getURL().toLowerCase();

if(filters.matcher(href).matches())
{
     return false;
     }

if(href.startsWith("http://www.ics.uci.edu/")) {
     return true;
     }
     return false;
}

public void visit(Page page)
{
     int docid = page.getWebURL().getDocid();
     String url = page.getWebURL().getURL();
     String text = page.getText();
     ArrayList<WebURL> links = page.getURLs();
     }
}
```

3.3. Re-visit Policy

In the proposed methodology, the parsers also use the re-visit policy when needed because web has its dynamic nature. The re-visit policy can be easily understood using the freshness function described in the following sub-sections.

3.3.1. Freshness

This is used as binary measure which indicates whether the local copy is accurate or not. The freshness of any page 'p' in the repository at time t is defined as:

$$F_p(t) = \begin{cases} 1 & \text{if } p \text{ is equal to the local copy at time } t \\ 0 & \text{otherwise} \end{cases}$$

3.3.2. Age

It is a measure which indicates how outdated the local copy of page is. The age of a page 'p' in the repository, at time t is defined as:

$$A_p(t) = \begin{cases} 0 & \text{if } p \text{ is not modified at time } t \\ t - \text{modification time of } p & \text{otherwise} \end{cases}$$

4. EXECUTION OF THE PROPOSED METHODOLOGY

Execution of proposed methodology depends on the sequence of knowledge bases and corresponding parsers. Final result of the proposed methodology is not affected by the sequence of the operations. Sequence affects only the space complexity and time complexity of the methodology.
Execution of proposed methodology in divided into several steps which are described as follows in the following sections.

4.1. Historical Attack Detection

This step is composed with 2 operations which are occurred using 'Knowledge Base I' and 'Knowledge Base T'.

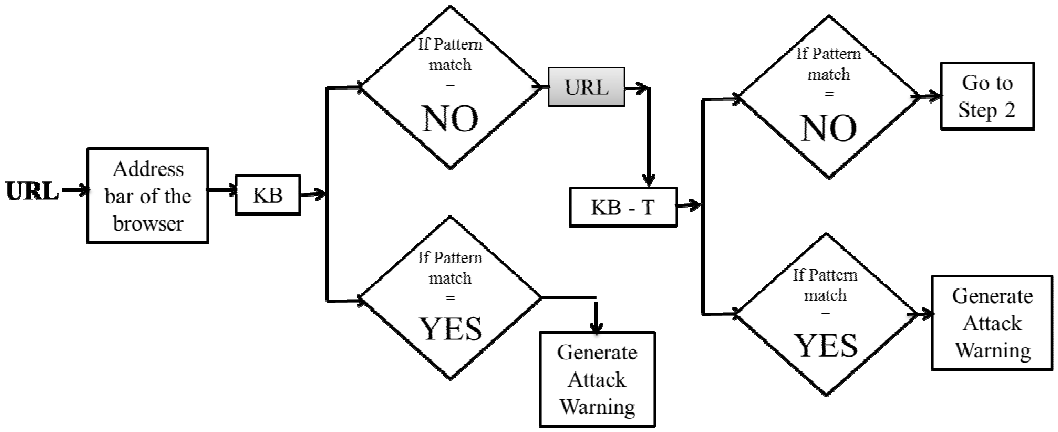


Figure 7. Flowchart of Step 1 of the proposed methodology

Knowledge Base Initial (KBI) is used to detect the attacks which has the same pattern with the previous detected attacks stored in it. Knowledge Base Trusted (KBT) is used to find the trusted status of requested URL which was previously declared by the user. In Historical attack detection the browser first tallies the URL with the KBI to check if its pattern matches that of any frequent phishing attack stored in the knowledge base. If it is safe then it proceeds to match up with the KBT. In this knowledge base it matches the URL against the trusted URLs stored by the user.

4.2. URL Pattern based Attack Detection

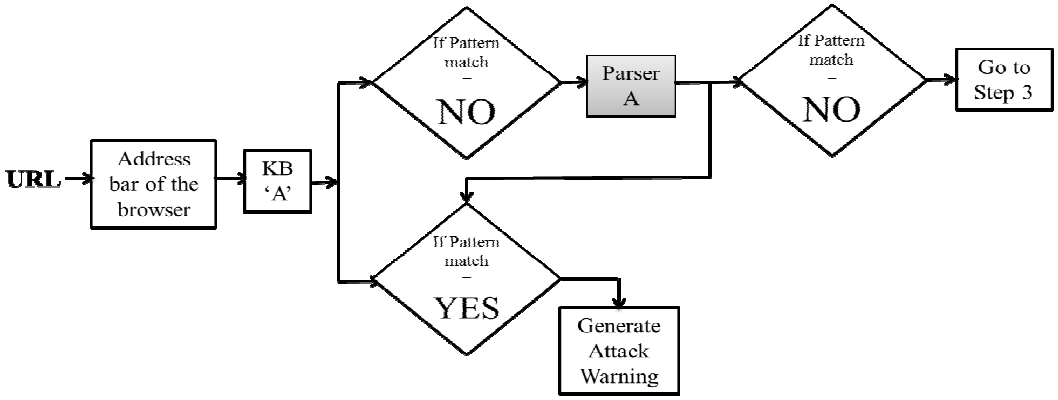


Figure 8. Flowchart of Step 2 of the proposed methodology

It is composed 2 operations which are related to 'Knowledge Base A' and 'parser 1'. This Step 2 provides the security against those attacks which are purely URL-pattern based phishing as well as SSL attacks. Knowledge Base A detects only those attacks which were detected previously by the browser and were stored in its database. During the step 2, 'parser 1' scans the requested URL and finds the occurrence of special characters ('-','.' etc) and their repetition in the URL. It is used to detect the pattern based phishing attacks. Generally phishing websites use these special characters repeatedly to hide its fraudulent nature. Working of step 2 is represented in Figure 8.
Google Chrome has auto SSL attack detection feature inbuilt within itself. Figure shows the Chrome behavior towards SSL URLs. In the browser, 'https' text is shown in red with an arrow sign to signify that the URLs being accessed have an invalid SSL certificate. Thus the page being accessed is encrypted but the license of the website has expired. Hence it can be a false website in place of the previous popular one which the user wanted to actually access.

4.3. URL Information Analysis

URL information can be very helpful in detection of the phishing attacks. This step is based on the fact that the phishing URLs are newly registered and have lower rating and popularity over the internet. Figure represents the working of URL information analysis step.

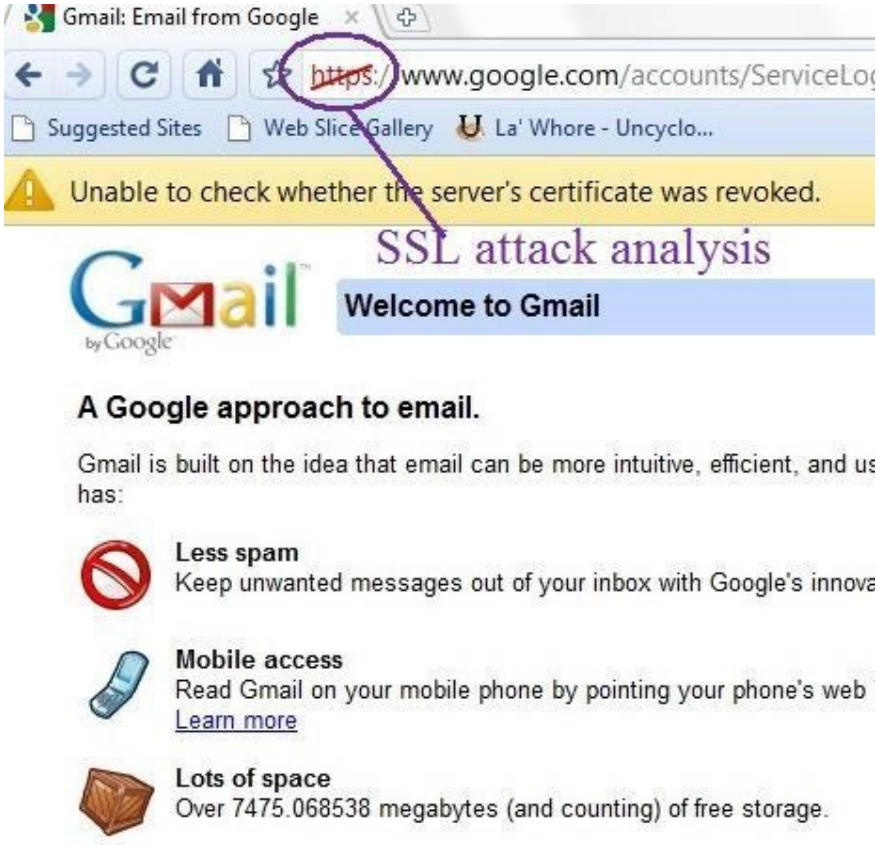


Figure 9. Representation of Gmail in Google Chrome

In this step requested URL is analysed with the Knowledge Base B and information of URL is analysed using the historical data of URL (if the URL was visited previously) and displays the results and generates warning if URL is phishing attack based URL.

If the URL is not present in the history of Knowledge Base B then it goes to the parser 2 for the information analysis. 'Parser 2' works to finds the information of URL as a web crawler (which is described above) and performs the proper analysis after crawling for the details of the URL over the internet. After all the details have been collected it generates a result depending on whether the URL is a popular site or a newly created one.

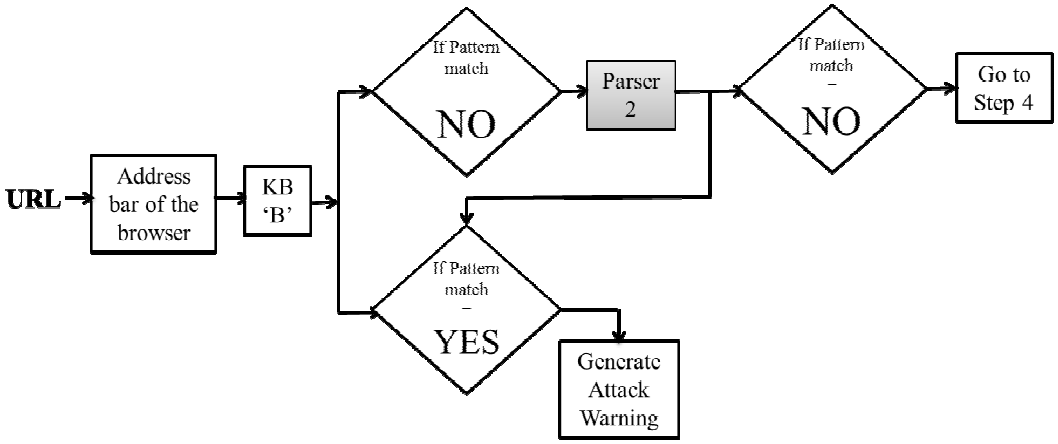


Figure 10. Flowchart of Step 3 of the proposed methodology

4.4. Fraud URL Detection

This step is performed by the Knowledge Base C and parser 3.Knowledge Base C performs the fraud check analysis of the requested URL (if it is available in the history of Knowledge base). It displays the result and appropriate messages. If the URL is not visited previously then parser 3 performs the Fraud check analysis to provide the security against phishing attacks (or other web attacks) using some security algorithms. Figure 13 describes the fraud check analysis.

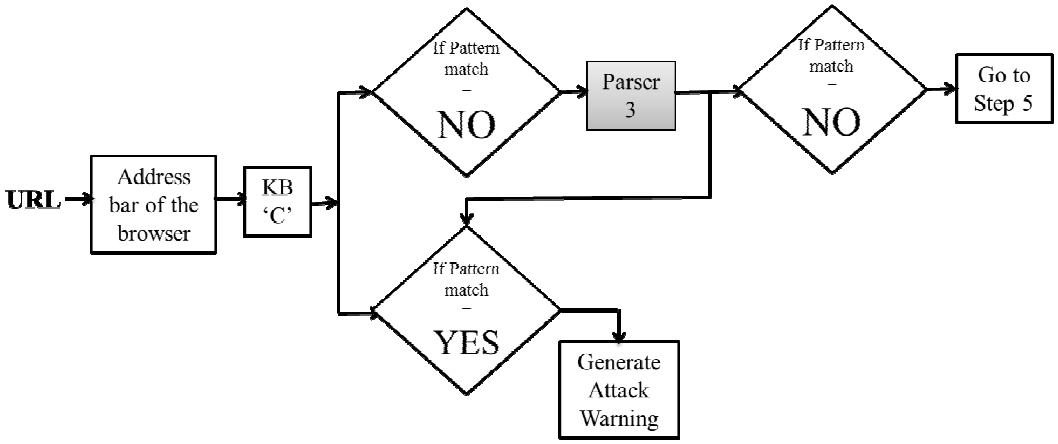


Figure 11. Flowchart of Step 4 of the proposed methodology

4.5. Comparison with other URLs

The above step is similar to the step 3, During the step 3, URL information analyzed using some assumption (phishing URLs have their early license year and lower rating level) of phishing attacks, but during the step 5, the URL information is compared with the other URLs information which have some similarities in URL string with the requested URL string.

Knowledge Base D provides the information of the requested URL and other URLs, using its history (if the URL is visited previously and history is maintained in the Knowledge Base) then compares all the information and produces the results.
If the URL is not visited previously, then the comparison is performed by the parser 4 using crawling operation over internet using some standard crawling techniques.

Figure 14 describes the step 5 of proposed methodology.

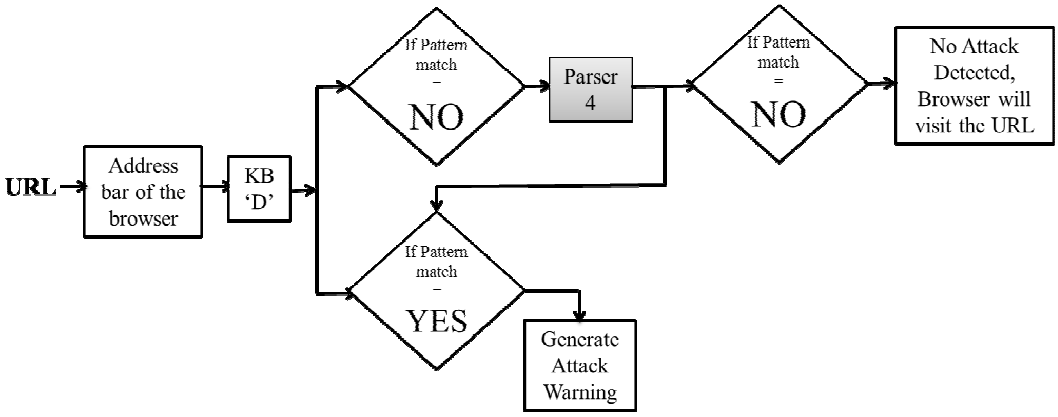


Figure 12. Flowchart of Step 5 of the proposed methodology

After completing the step 5, execution of proposed methodology will finish. Proposed methodology have more time complexity and space complexity but it provides the better security against web attacks in comparison with the other methodologies which are already proposed. The step by step approach ensures that a wide majority of web attacks are detected.

5. IMPLEMENTATION AND RESULTS

We have implemented the proposed methodology in a virtual scenario, where we explored all the visited URLs of browsers on different machines using the history feature. All the URLs have been stored in a database for detect the phishing attacks and perform the analysis. We have used Java programming, JSP and MySQL, apache tomcat web server to execute the proposed methodology. We have also implemented some advanced feature using build -network APIs and crawling. Proposed methodology also uses the crawling step to analyze the URL over World Wide Web. We are planning to implement this methodology with some new add-ons to install in present web browsers (like other Firefox add-ons) using some advanced techniques.

Table 1. URL and some web attacks analysis (2012 - 13)

Month	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**
No. of URLs visited	1098	1086	1007	1149	1368
Phishing Attacks	24	20	19	25	27
Detected phishing attacks with the browser	17	15	17	22	27
SSL Attacks	21	15	14	12	15
Detected SSL attacks with the browser	16	13	13	11	15
Execution Time (in minutes)	0	0	161	202	281

We have analysed the URL visited over the 5-months of period. In starting stage of implementation, security risks are more because of absence of data in the different knowledge

base. The implemented scenario provides 98.14 % security against phishing attacks and some hacking attacks. We have not executed our proposed methodology for the duration of Oct, 2012 and Feb, 2013 but during Dec, 2012 to Feb, 2013, we have executed the above methodology.
The above table shows the number of phishing attacks encountered and the execution time taken by our methodology from October 2012 to Feb 2013. The execution time for the first two months is actually zero as we have not implemented our methodology then. We have implemented our methodology from December 2012 onwards.

Kindly note that the approximate time of execution per URL visit, for the first month comes out to be about 11 seconds. This increases to 12 seconds in the second month and to 14 seconds in the third month. This gradual increase can be attributed to the fact that the knowledge base is increasing in size hence the browser searches for more security attack then before.

3. CONCLUSION AND LIMITATIONS

Our proposed methodology is inspired by a problem with a large number of Phishing, SSL and other web attacks, we have encountered. We have recorded the web URLs activities of with the usage of proposed methodology and without usage of proposed methodology over 5 months. From data, we have analysed the attacks and detected attacks over the time. The experiment results provide the complete scenario of the problem and security over the web. Our system indicated that the 96.94% security against phishing attacks as well as SSL-attacks over the browsing. Table 1 represents the recorded data over the 5 months' time period.

Limitations of the proposed method are that due to various parsing operations, its time complexity and space complexity is higher. So many times, it increases the browsing time of web browser. Due to slower speed of browsing, generally web users avoid this type of higher web security.

REFERENCES

[1] PHP, AJAX, MySQL and JavaScript Tutorials, http://www.w3schools.com/

[2] Prentice Hall - Deitel - Java How to Program, 4th Edition, Java_2_Complete_Reference_5E, Java - How to Program, 6th Edition

[3] Ahn, Luis von, Blum, Manuel, Hopper, Nicholas, and Langford, John. CAPTCHA: Using Hard AI Problems for Security. In Eurocrypt

[4] Gedam,Dhiraj Nilkanthrao,RSA Based Confidentiality And Integrity Enhancements in SCOSTA-CL, A thesis report, Department of Computer Science and engineering, Indian Institute of Technology ,Kanpur, India, (July,2009)

[5] J. Ullman, Database and knowledge base systems, In Database and knowledge-base systems, Volume 2, Computer Science Press, 1989

[6] Akerkar RA and Sajja Priti Srinivas: "Knowledge-based systems", Jones & Bartlett Publishers, Sudbury, MA, USA (2009)

[7] Delayed password disclosure Markus Jakobsson, Steven Myers, Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94303, USA. School of Informatics, Indiana University, Bloomington, IN, USA Journal: Int. J. of Applied Cryptography 2008 - Vol. 1, No.1 pp. 47 - 59

[8] Herzberg, A. and Gbara, A. (2004) Technical Report 2004-23, Protecting (even) Naïve Web Users, or: Preventing Spoofing and Establishing Credentials of Web Sites, DIMACS, October 30, Available at: http://dimacs.rutgers.edu/TechnicalReports/2004.html., S.hyun. & Kim Mi Na, (2008) "This is my paper", ABC *Transactions on ECE*, Vol. 10, No. 5, pp120-122.

[9] Sophos White Paper, Phishing and the threat to corporate networks, (2005)

[10] Beginning PHP5, Apache, and MySQL Web Development by Elizabeth Naramore, Jason Gerner, Yann Le Scouarnec, Jeremy Stolz, Michael K. Glass; ISBN: 9780764579660

[11] Chen, Juan and Guo, Chuanxiong, Online Detection and Prevention of Phishing Attacks, in Proc. Chinacom '06.

[12] Alnajim, A. and Munro, M. 2009. An Anti-Phishing Approach that Uses Training Intervention for Phishing Websites Detection. In Proceedings of the 2009 Sixth international Conference on information Technology: New Generations (2009). ITNG. IEEE Computer Society, Washington, DC, 405-410. DOI= http://dx.doi.org/10.1109/ITNG.2009.109

[13] Yu, W.D.; Nargundkar, S.; Tiruthani, N., "A phishing vulnerability analysis of web based systems," Computers and Communications, 2008. ISCC 2008. IEEE Symposium on, vol., no., pp.326-331, 6-9 (July 2008)

[14] Abu-Nimeh, S.; Nair, S., "Bypassing Security Toolbars and Phishing Filters via DNS Poisoning," Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE, vol., no., pp.1-6, (Nov. 30 2008-Dec. 4 2008)

[15] Aburrous, Maher Ragheb, Alamgir, Hossain,Keshav Dahal, Thabatah, Fadi, "Modelling Intelligent Phishing Detection System for E-banking Using Fuzzy Data Mining," cw, pp.265-272, 2009 International Conference on CyberWorlds, (2009)

[16] Ollmann G., the Phishing Guide Understanding & Preventing Phishing Attacks, NGS Software Insight Security Research

Authors

Gaurav Kumar Tak is an assistant professor in the School of Computer Engineering, Lovely Professional University. His primary research areas of interest are Cyber-crime and Security, Wireless Ad-hoc Network and Web Technologies. He has written several research papers in these areas and continues to work for improving the security of web applications and making the web safe to surf.

Gaurav Ojha is a student in the Department of Information Technology, Indian Institute of Information Technology and Management, Gwalior. His areas of interest are Web technologies, Open source software and Internet Security.