JCSE International Journal of Computer Sciences and Engineering Open Access

Study Paper Volume-4, Issue-1 E-ISSN: 2347-2693

A Study on Anti-Phishing Techniques

V. Raghunatha Reddy[1], C. V. Madhusudan Reddy[2], M. Ebenezar[3*]

[1] *Department of Computer Science and Engineering, SKU, Anantapur, India,*
[2,3*] *Department of Computer Science and Engineering, SJCET, JNTU Anantapur, India,*

www.ijcseonline.org

Received: Dec/02/2015 Revised: Dec/14/2015 Accepted: Jan/11/2016 Published: 30/Jan/2016

Abstract— Some customers avoid online banking as they perceive it as being too vulnerable to fraud. The security measures employed by most banks are never 100% safe; Credit card fraud, signature forgery and identity theft are far more widespread "offline" crimes than malicious hacking. An increasingly popular criminal practice to gain access to a user's finances is phishing, whereby the user is in some way persuaded to hand over their password(s) to the fraudster. To protect users against phishing, various anti-phishing techniques have been proposed that follows different strategies like client side and server side protection. In general anti-phishing techniques are Content Filtering, Black Listing, Symptom-Based Prevention, Domain binding, Character based Anti-Phishing, Content based Anti-Phishing. But they have got some drawbacks such as; Time Delay, Redundancy, Accuracy, Information Retrieval.The Proposed system will check the user’s existence in the database and provide the set of services with respect to the role of the user. The application is based on three-tier architecture. The cipher key will be used to find the fraud application. This approach is called Anti-Phishing. Anti-Phishing is nothing but “preventing the phishing”. This can be done by creating a cipher key (an encrypted code) in the customer’s username, password or in a/c no., which is not recognized in the hacker’s fake website. The objective of the project is to design and develop secure online Banking Application using Anti-phishing concept.

***Keywords*—**Anti-Phishing, Trojan, Black List, Proxy, Spyware, Cipher Key, Spoofguard

I. INTRODUCTION

Innovation with information and communications technologies (ICTs) has permeated our lives, be it for business, for personal or for recreation purposes. We especially rely on the Internet for business, personal, finance and investment decision making, etc. Coupled with these advancements are myriad of threats which exploit the inherent vulnerabilities on the Internet and its associated technologies. Some of these threats manifest in various ways, such as pretentious items offered for sale on eBay, with the aim to swindle unsuspecting patrons, or assuring victims of great returns, if the victim will help a foreign financial transaction through his own bank account, etc. Phishing is a fraudulent attempt to gain personal information from victims such as bank information, credit card information, social security, employment details, and online shopping account passwords and so on. Phishing attacks use fraudulent e-mails or websites designed to fool users into divulging personal financial data by stealing the trusted brands of well-known banks, e-commerce and credit card companies. People regularly trust about any information they receive through email or website and the attackers use injection attacks to hide his website by email or by URL redirection. Phishing websites are like legitimate websites, even to the point of using the graphics and links straight of the legitimate website.

While phishing tricks are continuously growing, one common trick is to have a login screen in a popup window, which allows them to copy the legitimate website exactly. Attackers send an email contain a hyperlink to open a new window or popup windows while browsing the web that claims to be legitimate website. The popup window may ask to update, validate or confirm account information and it is like official organizations websites [1]. One of the W3C standards is avoid using popup windows or using a hyperlink to open in a new window in the webpage. So, we conclude that attackers use some tricks to fool the user and tempting them, some of these are external links for images or logos, suspicious URLs, external domains, email, iframe, suspicious script, and popup window [2].

The article is structured as follows: The next section provides a brief overview of the different types of phishing attacks. Section 3 describes Classic Attack Vectors, Section 4 presents an overview on Anti-phishing Techniques. Section 5 discusses related work. Section 6 presents future work and Section 7 concludes the article.

II. TYPES OF PHISHING ATTACKS

In this section, we give a brief overview of the different types of phishing attacks to familiarize the reader with the threat.

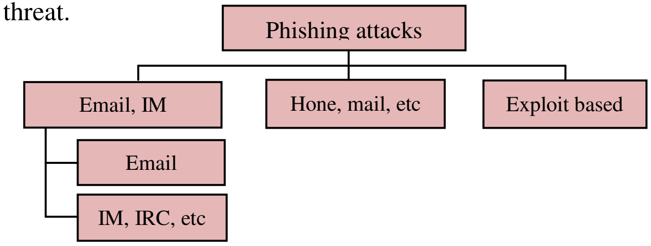

Figure1. Phishing attacks

The Figure 1 illustrates various ways how phishing attacks may be done.

II.1. Spoofing e-mails and web sites

Phishing attacks fall into several categories. The earliest form of phishing attacks were e-mail-based and they date back to the mid 90’s. These attacks involved spoofed e-mails1 that were sent to users where attackers tried to persuade the victims to send back their passwords and account information. Although such attacks may be successful today, the success rate from the point of view of the attackers is lower because many users have learned not to send sensitive information via e-mail. A possible reason is that many security- sensitive organizations such as banks do not provide interactive services based on e-mail where the user has to provide a password. Most organizations, obviously, use their web sites for providing interactive services because they can rely on encryption technologies such as SSL. As a result, a typical user would find a request to send sensitive information such as a password via e-mail suspicious (especially considering the fact that many Internet users today receive a large number of spam e-mails from people that they do not know).

Hence, many phishing attacks now rely on a more sophisticated combination of spoofed e-mails and websites to steal information from victims. Such attacks are the most common form of phishing attacks today.

In a typical attack, the attackers send a large number of spoofed e-mails that appear to be coming from a legitimate organization such as a bank to random users and urge them to update their personal information.

The victims are then directed to a web site that is under the control of the attacker. This site looks and feels like the familiar online banking web site and users are asked to enter their personal information. Because the victims are directly interacting with a web site that they believe one Taking advantage of the weaknesses in the SMTP protocol, attackers can easily fake the From e-mail header and spoof e- mails. They know, the success rates of such attacks are much higher than e-mail-only phishing attempts [3].

Besides e-mail, as an alternative form of message delivery, attackers have also started to use instant messaging systems such as ICQ or infrastructures such as Internet Relay Chat (IRC) to try to persuade and direct users to spoofed web sites.

Once the victim follows a spoofed link, in order not to raise suspicion and to present the phishing web site as authentic as possible, attackers are employing various techniques. One example is the use of URLs and host names that are obfuscated and modeled so that they look legitimate to inexperienced users. Another example is the use of real logos and corporate identity elements from the legitimate web site. Some attacks also make use of hidden frames and images as well as Java-script code to control the way the page is rendered by the victim’s browser.

II.2. Exploit-based phishing attacks

Some phishing attacks are technically more sophisticated and make use of well-known vulnerabilities in popular web browsers such the Internet Explorer to install malicious software (i.e., malware) that collects sensitive information about the victim. A key logger, for example, might be installed that logs all pressed keys whenever a user visits a certain online banking web site. Another possibility for the attacker could be to change the proxy settings of the user’s browser so that all web traffic that the user initiates passes through the attacker’s server (to perform a typical man-in-the- middle attack).

Exploit-based phishing attacks are not the focus of our work in this article. To mitigate exploit-based phishing attacks (as well as other security threats that are directly related to browser security such as worms, trojans and spyware), browser manufacturers need to make sure that their software is bug-free and that users are up to date on the latest security fixes. The focus of Anti-Phishing is to mitigate web site-based phishing attacks that aim to trick victims into giving away their sensitive information.

III. CLASSIC ATTACK VECTORS

III.1. Social Engineering

Social Engineers are experts at appealing to the human psyche. Their most common methods of manipulation rely on Curiosity, Fear, and Empathy. A Social engineer uses these tactics in phishing scams. These tactics are described below:

- Curiosity. Exploiting a person's curiosity might involve sending an e-mail that purportedly contains a link to watch a video about the latest sensational news story. The link, however, will lead to a malicious site aimed at installing malware or stealing private information.
- Fear. One tactic cyber thieves use to instill fear and persuade a person to act in a certain way is by sending phishing e-mails, supposedly from a victim's bank. Using the claim that his or her account has been breached, the message will push the user to click a certain link to validate the account. Again, the link will

lead to a malicious site aimed at compromising the person's computer, or stealing sensitive information.

- Empathy. To take advantage of a person's empathetic feelings towards others, hackers have been known to impersonate victims' friends on networking sites, claiming to urgently need money. In another prime example, recent social engineering scams have also been seen in the wake of the earthquake and tsunami in Japan, with scammers attempting to profit from the tragedy [4].

III.2. Trojaned Hosts

Gifts are a common tactic that attackers use as bait to catch their phish. The ultimate prize of a attacker is a trojaned host. An attacker will send a phishing email that offers the victim some type of gift if the victim visits a website. The gift that the victim receives is a Trojan horse program [5] on their computer. Typically, the Trojan program allows the attacker to have remote control of the computer and access to all information on the computer and information that the computer processes. One of the greatest uses of the compromised computer is propagating the attacker's spam messages.

Unfortunately there is a Trojan creation toolkit available called Zeus (keeping with the mythological theme). Zeus, which is sold on the black market, allowing non-programmers to purchase the technology they need to carry out cybercrimes. According to a 2010 report from SecureWorks, the basic Zeus package starts at about $3,000. The scary thing about the Zeus package is that a study found "most of the infections occurred on machines where an antivirus product was installed and kept up-to-date: 31% of the Zeus-infected PCs had no antivirus while 55% had updated antivirus software". Zeus Trojans are typically not detected by anti-virus and is considered a zero-day exploit due to it being a custom tool for creating unique Trojans that are not "in the wild". When the Trojan is created, if the creator limits the use of it, the anti-virus vendors will never add the signature to their databases due to its low distribution.

III.3. Man-in-the-middle Attacks

A Man-in-the-Middle (MitM) Attack is the result of an attacker inserting themselves logically into the network between a victim and a legitimate website.

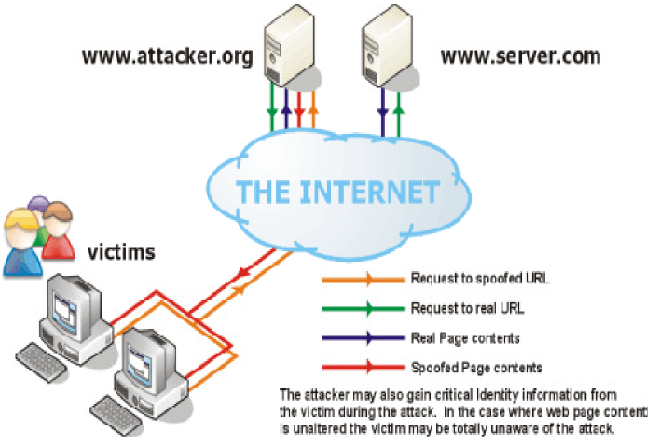
www.attacker.org
www.server.com
THE INTERNET
victims
Request to spoofed URL
Request to real URL
Real Page contents
Spoofed Page contents
The attacker may also gain critical Identity information from the victim during the attack. In the case where web page content is unaltered the victim may be totally unaware of the attack.

Figure2. Man in the Middle

A MitM attack employs a number of attack methods such as transparent proxies, DNS Cache Poisoning, URL Obfuscation, and browser proxy configuration attacks.

III.3.1. Transparent Proxies

This type of attack uses a traditional phishing methodology of sending a deceptive email to an intended victim. Contained in this email is a spoofed URL for a targeted website. The attacker's intention is to steal personal information. In the figure below the attacker, noted as attacker.org, acts as proxy for the entire transaction. The victim is completely unaware of the proxy in this attack.

In this situation, the attacker copies the entire source code from the original website. The attacker modifies the original HTML code and uploads it to the spoofed website. The spoofed website interacts with the source website on behalf of the victim. Following the network lines in the figure2, the victim never interacts with the source website directly and the attacker can gain all of the victim's information.

III.3.2. DNS Cache Poisoning.

Domain Naming System (DNS) Cache Poisoning is an almost undetectable phishing method. "DNS cache poisoning, is the corruption of an Internet server's domain name system table by replacing an Internet address with that of another, rogue address".

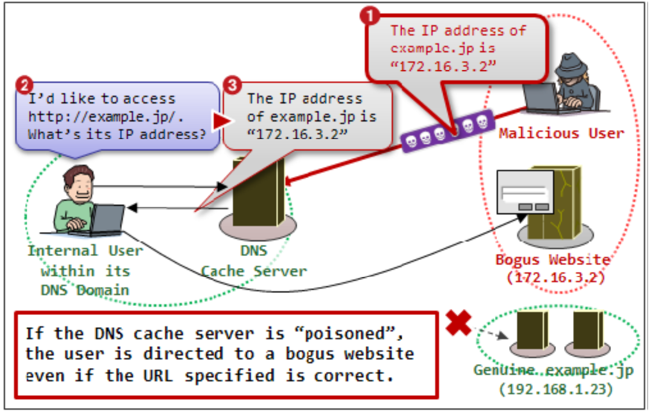
1
The IP address of cxamplc.jp is "172.16.3.2"
2
I'd like to access http://example.jp/. What's its IP address?
3
The IP address of example.jp is "172.16.3.2"
Malicious User
Internal User within its DNS Domain
DNS Cache Server
Bogus Website (172.16.3.2)
If the DNS cache server is "poisoned", the user is directed to a bogus website even if the URL specified is correct.
Genuine example.jp (192.168.1.23)

Figure3. DNS Cache Poisoning [6]

The victim in this case does not respond to an unsolicited email. The victim enters a legitimate web address into a browser. The attacker altered the DNS lookup tables on the DNS server. Web addressing is Internet Protocol (IP) number based. IP addresses in version 4 are based on 4 octets. IP addresses follow this format xxx.xxx.xxx.xxx. Remembering an IP address is difficult for users so DNS was invented so that users can use common names. The computer resolves the IP from a DNS server. Resolving means that common names are automatically converted to the IP format. The DNS server has look up tables to make name resolution faster. In DNS poisoning the attacker changes the IP address that is associated with the common name to match a bogus website that the attacker has set up. Figure 3 above outlines the process.

IV. ANTI-PHISHING TECHNIQUES

Anti-phishing defenses can be server and client based solutions.

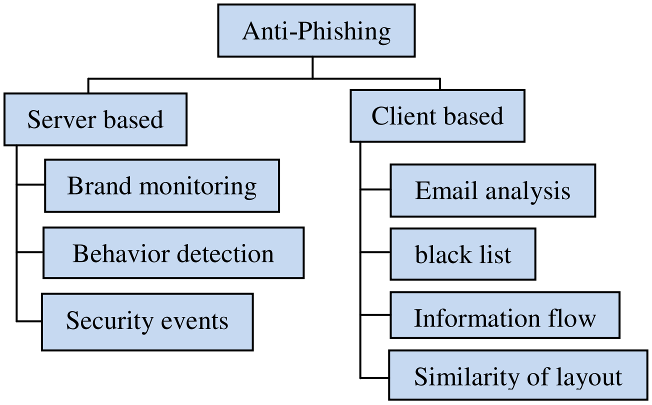

Figure4. Anti-Phishing Techniques

Server Based- these techniques are implemented by service providers (ISP, etc) and are of following types:

- ➢ Brand Monitoring: Cloning online websites to identify "clones" which are considered phishing pages. Suspected websites are added to centralized "black list".
- ➢ Behavior Detection: for each customer, a profile is identified (after a training period) which is used to detect anomalies in the behavior of users.
- ➢ Security Event Monitoring: security event analysis and correlation using registered events provided by several sources (OS, application, network device) to identify anomalous activity or for post mortem analysis following an attack or a fraud.

Client Based-these techniques are implemented on user's end point through browser plug-ins or email clients and are of following types:

- ➢ Email based analysis: email based approaches typically use filters and content analysis. If trained regularly, Bayesian filters are actually quite effective in intercepting both spamming and phishing e-mails. Bayesian algorithm explains the working of Bayesian filter:

Bayesian Algorithm:

1) Split e-mail in tokens.
    - Need number of messages for spam and legitimate.
    - Need frequency of each word for each type.

2) Calculate probabilities.
    - P (legitimate) = word frequency /number of legitimate messages.
    - P (spam) = word frequency/ number of spam messages.

3) Calculate likelihood of being spam (spamicity) using a special form of Bayes' Rule where likelihood = a/(a + b), where a is the probability of a legitimate word and b is the probability of spam word.
4) Choose tokens whose combine probability is farthest from 0.5 either way. This is because the farther it is from 0.5 (neutral), with more certainty we can say it belongs to either strategy.
    - Do this for n numbers for n instance
    - Combine their probability to get a figure for message using Bayes' Rule. In basic terms, Baye's Rule determines the probability of an event occurring based on the probabilities of two or more independent evidentiary events. For three evidentiary events a, b, and c, the probability is equal to

$$\frac{a\ b\ c}{abc+ (1-a)*(1-b)*(1-c)}$$

    - If the end result is closer to 1.0, then the message is classified as spam, and if it is closer to 0.0, the message is classified as legitimate.

- ➢ Black Lists: black lists are collection of url's identified as malicious. The black-list is queried by the browser at run time whenever a page is loaded. If the currently visited url is included in the black list, the user is advised of the danger otherwise the page is considered legitimate.

V. RELATED WORK

Phishing website is a huge effect on the financial and online commerce, detecting and preventing this attack is an important step towards protecting against website phishing attacks, there are several approaches to detect these attacks. In this section, we review existing anti phishing solutions and list of the related works.

One approach is an intelligent Phishing Website Detection System using Fuzzy Techniques [7]. It is based on fuzzy logic and produces six criteria's of website phishing attack. There are many characteristics and factors that can distinguish the original legitimate website from the forged faked phishing website like spelling errors, long URL address and abnormal DNS record. Website phishing detection rate is performed based on six criteria and there are different numbers of components for each criterion, the criteria are:

1- URL & Domain Identity.
- a) Using the IP address.
- b) Abnormal request URL.
- c) Abnormal URL of anchor.
- d) Abnormal DNS record.
- e) Abnormal URL.

2- Security & Encryption.
- a) Using SSL certificate.
- b) Certification authority.
- c) Abnormal cookie.
- d) Distinguished Names Certificate (DNC).

3- Source Code & Java script.
- a) Redirect pages.
- b) Straddling attack.
- c) Pharming attack.
- d) Using onMouseOver to hide the Link.
- e) Server Form Handler (SFH).

4- Page Style & Contents.
- a) Spelling errors.
- b) Copying website.
- c) Using forms with "Submit" button.
- d) Using Pop-up windows.
- e) Disabling right click.

5- Web Address Bar
- a) Long URL address.
- b) Replacing similar characters for URL.
- c) Adding a prefix or suffix.
- d) Using @ symbol to confuse.
- e) Using hexadecimal character codes.

6- Social Human Factor.
- a) Much emphasis on security and response.
- b) Public generic salutation.
- c) Buying Time to Access Accounts.

The rule base has input parameters (criterion) and one output that contain all the "IF-THEN" rules of the system. The output for each criterion is one of the following: Genuine, Doubtful or Fraud. The output of final website phishing is one of the final output (Very Legitimate, Legitimate, Suspicious, Phishy or Very Phishy) which representing final phishing website rates.

Second approach is a client-side defense against web-based identity theft. It proposes a framework for client-side defense: a browser plug-in called SpoofGuard that examines webpages and warns the user when requests for data may be part of a spoof attack, it computes a spoof index (a measure of the likelihood that a specific page is part of a spoof attack), and warns the user if the index exceeds a level selected by the user. SpoofGuard uses a combination of page evaluation and examination of outgoing post data to compute a spoof index.

When a user enters a username and password on a spoof website that contains some combination of suspicious URL, misleading domain name, images from an honest website, and a username and password that have previously been used at an honest website, SpoofGuard will intercept the post and warn the user with a popup that foils the attack. The browser plug-in applies tests to all downloaded pages and combines the results using a scoring mechanism. The total spoof index of a page determines whether the plug-in alerts the user and determines the severity and type of alert. Since popup warnings are intrusive and annoying, it attempts to warn the user through a passive toolbar indicator in most situations.

In order to apply image and URL check, the SpoofGuard plug-in is supplied with a fixed database of images and their associated domains. When the browser downloads a login page all images on the page are compared to images in the SpoofGuard database. The spoof-score for the page is increased if a match is found but the page's domain is not a valid domain for the image. The browser history file and additional history stored by SpoofGuard are used to evaluate the referring page. When a user fills in form data, SpoofGuard intercepts and checks the HTML post data, allowing the actual post to proceed only if the spoof index is below the user specific threshold for posts.

Third approach is Anomaly Based Web Phishing Page Detection [8]. It examines the anomalies in webpages, in particular, the discrepancy between a website's identity and its structural features and HTTP transactions. A structured webpage is composed of W3C DOM objects. Among them, it lists five categories, based on their relevance to the web identity. They are Keyword/Description (KD), Request URL (RURL), URL of Anchor (AURL), Server Form

Handler (SFH) and Main Body (MB).These categories are the main sources which the identity and features are derived from and lists the characteristics of phishing like Abnormal URL, Abnormal DNS record, Abnormal Anchors, Abnormal Server Form Handler, Abnormal Request URL, Abnormal cookie and abnormal certificate in SSL.

It extracts the related web objects from a webpage and converts them into a feature vector based on the characteristics of phishing analysis. The page classifier takes the feature vector as input and determines whether the page is bogus or not. The proposed phishing detector consists of two components Identity Extractor and Page Classifier.

Identity Extractor uniquely identifies the website's ownership; the identity is an abbreviation of the organization's full name and/or a unique string appearing in its domain name.

Page classifier refers to these objects/properties as structural features. One source of structural features is those identity related W3C DOM objects in a webpage, e.g. URI domain of an anchor. Another source of structural features is HTTP transactions. Page classifier employs Support Vector Machine (SVM), a well-known algorithm for classification. It outputs a label 1 which indicating a phishing page or a label -1 which indicating an authentic one.

To facilitate SVM based classification, they quantify those features into vectors. The output from the execution of the webpage identity extractor is a character strings from extracted identity words. The feature vector initialization of webpage are URL address, DNS record ,URL of anchor , request URL , server form handler , domain in cookie and certificate in SSL. Given an identity and a set of features, the task of determining the genuineness of a webpage is executed by Support Vector Machine (SVM), which is a well-known classifier and has been widely employed in pattern recognition.

From the approaches we conclude that, there are many characteristics and anomalies can be found in the webpage and we can detects any possible attacks based on these characteristics, attackers use these characteristics in phishing webpages to gain sensitive information from users.

VI. Future Work

We are currently working on the changing the implementation of online banking websites. It is because most of the phishing attacks happen through spoofing online banking websites.

We are working towards facilitating the online banking websites with an ability to create a cipher key from the user credentials present in the database, the cipher key thus generated can be further used for the website authentication. The fake websites cannot authenticate the user as the original websites have a different mechanism.

VII. Conclusion

Phishing it is a kind of online identity theft. The main aim of the attackers is to steal the sensitive information from users such as online banking passwords and credit card information. The recent years have brought a drastic increase in the number and sophistication of such attacks. Though phishing scams have received extensive press coverage, phishing attacks are still successful because of many inexperienced and unsophisticated Internet users. Attackers are employing a large number of technical spoofing tricks such as URL obfuscation and hidden elements to make a phishing web site look authentic to the victims.

The attackers are trying to attack with a profound technical knowledge. There is an immense need of new anti-phishing tools and plug-ins as the phishing attacks are increasing day-by-day.

The most effective solution to phishing is training the users not to blindly follow the links to web sites where they are asked to enter their sensitive information such as passwords, date-of-birth, bank account numbers. There will always be users that are tricked into visiting a phishing web site. Therefore, it is more important for the users to have a brief idea on the phishing attacks and the way they happen.

References

[1] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, and J. C. Mitchell, "Client–side defense against web–based identity theft", In Proceedings of 11th Annual Network and Distributed System Security Symposium, 2004.

[2] Popup Window, "http://www.w3.org/TR/WAI-WEBCONTENT/", W3C Recommendation 5-May-1999

[3] Ollman, G. (2004), "The Phishing Guide – Understanding and Preventing Phishing Attacks", IBM Internet Security Systems.

[4] Atkins, B. and Huang, W. (2013), "A Study of Social Engineering in Online Frauds", Open Journal of Social Sciences, 1, 23-32. doi: 10.4236/jss.2013.13004.

[5] Trojan horse program, "http://searchsecurity.techtarget.com/definition/Zeus-Trojan-Zbot", 2010

[6] DNS Cache Poisoning, "http://www.ipa.go.jp/security/english/vuln/200809_DNS_en.html", Sep 18, 2008

[7] M. Aburrous, M.A. Hossain, F. Thabatah and K. Dahal, "Intelligent phishing website detection system using

fuzzy techniques", in 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA), pp. 1-6, 2008.

[8] Y. Pan and X. Ding, "Anomaly Based Web Phishing Page Detection", Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06), Computer Society, 2006.

AUTHORS PROFILE

Dr. V. Raghunatha Reddy obtained his B. sc From Sri Krishnadevaraya University, Ananthapur, India in 1991 and MCA from Madurai kamaraju University, Tamilnadu India in 2002, M. Phil from MKU Tamilnadu India in 2005 and Ph. D degree from SKU Ananthapur, India in 2009.
He published two international paper in reputed journal and attended three National Conference. He is at present working as Assistant Professor in the Department of Computer Science Sri Krishnadevaraya University, Ananthapur, Andhra Pradesh, India.

C. V. Madhusudan Reddy obtained his B.Tech From St. John'sCollege of Engineering and Technology, Yemmiganur, India in 2006 and M. Tech from KVN College of Engineering, Gulbarga University, Karnataka, India in 2012. He is working as Assistant Professor in the Department of Computer Science, St. John's College of Engineering and Technology, Yemmiganur, Andhra Pradesh, India.

M. Ebenezar was born in Andhra Pradesh, India. He obtained B. Tech degree at the R.G.M College of Engineering and Technology, Nandyal and he is currently Persuing his M. Tech degree in Computer Science and Engineering from St. John's college of Engineering and Technology, Yemmiganur.