ALITERATURE SURVEY OF VISUAL SIMILARITY SNOOPING ATTACKS IN EMAILS

George Mwangi Muhindi[1], Dr. Geoffrey Mariga Wambugu[2], Dr. Aaron Mogeni Oirere[3]

[1,2,3]School of Computing and Information Technology, Murang' a University of Technology,

Kenya

ABSTRACT

Snooping is one of the biggest issues that the cyber security industry is facing in this modern era of technology and it leads to big losses of finances both for individuals and organizations. The detection of snooping attacks with efficiency and preciseness is proving to be a challenge due to the complex nature of the snooping attacks. A snooping website appears to be very similar to the corresponding genuine website which deceives the unknowing users to believing that they are on the correct site. Banks and other financial institutions should prevent loss of money through snooping attacks. To achieve this, they should understand how the snooping attacks occur and the techniques that can be used to detect visual similarity snooping attacks. There is also need to develop and implement a mechanism that can check against snooping attacks and this can be achieved by checking for malicious links and attachments.

Keywords

Snooping Attacks, Websites, Email, Security, Malware

1. INTRODUCTION

The email security threat has risen to become one of the biggest threats to companies across the world. Subsequently, it can be noted that a majority of the hacking attacks begin with some form of snooping attack. Snooping can be described as a kind of attack which is engineered socially to steal private and confidential data like passwords, credit card information and login details. These snooping attacks take place when the attackers masquerade as genuine and trusted entities and end up tricking the unknowing users into opening the spammed emails [1]. When the recipient of the email receives the email and clicks on the embedded emails, malicious malware from the links infiltrate the computer system and, in the process, end up accessing and stealing private, sensitive and confidential information of the user.

The fake emails normally look very legit and genuine and even the links that the user is asked to click on appear to be very legit when they request for personal information. The snooping messages propagate themselves past the instant messages, social media sites, emails and VoIP. Nonetheless, email is the most popular way of carrying out the snooping attacks. This is true since 65% of the snooping attacks take place when the user clicks on a link and visits the hyperlink that is attached in the snooped email. More complicated snooping attacks target specific persons or groups from a firm [2].Metaphorically, snooping is the same as fishing in a lake, rather than attempting to fish a fish, the attackers attempt to steal the personal information of the user [3]. When the user unknowingly opens the fake website and feeds his or her personal information such as login details, these personal details are acquired by the hacker who can then use this information for other malicious intentions.

The snooping websites have an appearance that is very similar to the genuine website in order to attract a larger number of users to the website. With the development of snooping detection techniques, new approaches have been developed to detect visual similarity attacks [4]. Visual similarity-based techniques use comparisons of the visual appearance of the websites that are suspicious in correspondence to the genuine website by analyzing different parameters.

Banks should ensure that they prevent financial loses that may occur as a result of snooping attacks and should come up with techniques of preventing more snooping attacks. Moreover, there should be a technique that should check for snooping attacks before they occur. This should be done by checking for malicious links and attachments in the snooped emails that are usually sent toun suspecting victims [5] [4].

2. BACKGROUND AND STATISTICS OF SNOOPING ATTACKS

Snooping attacks and scams have gained the attention of both corporate and academic scholars since this issue has led to serious privacy breaches and adverse security issues in the banking industry resulting to loss of millions of dollars. Snooping attacks cannot be mitigated through the use encryption software and firewalls.

The first snooping attacks that were experienced took place on the American online network systems (AOL) which occurred during the onset of the 1990s. There were many fraudulent users that were registered on the AOL site using fake credentials and the AOL verified fake accounts using a simple test without analyzing the validity of the credit cards. After activating the fake accounts, the hackers were able to access the different resources that were offered by the American online system. During the billing process, AOL was able to find out that the accounts were illegitimate together with the fact that the linked credit cards were not valid [6]. Thus, the AOL stopped and closed down the accounts with immediate effect. After this incident, the American online networks system put in place measures that would ensure that the same does not happen in the future. The AOL put in place measures to prevent this through verifying and authenticating the credit cards that were linked to the billing accounts. This also enabled the attackers to switch, making it possible for them to obtain the AOL accounts. Rather than now creating the fake accounts, they changed to stealing the personal data of the users that were registered on the AOL system. The attackers then went ahead and contacted the registered users using emails and instant messages, requesting them to verify their personal information and passwords for security reasons [7]. The emails and the messages appeared to be originating from the AOL employees and this ended up duping a majority of the users to providing their personal information and passwords to the hackers. The attackers in turn used personal information in place of the valid customers. Here, the attackers did not restrict themselves to masquerading as the actual AOL users but also, they actively tricked many other commercial websites in the USA.

As per research by the Internet World Stats, the total number of users on the internet stood at 2.97billion in 2014 and by 2019 this number stood at 4.39 billion. This number is expected to rise up as years keep on going. With these figures in mind, over 38% of the global population makes use of the internet [8]. The large number of internet users gives hackers the chance of taking advantage of unknowing users and the insecure online systems to scam the users. Snooping emails are used for defrauding people and financial firms of money using the internet [9].

In the year 2012, there was a general increase in the number of snooping attacks translating to a 160% increase from the previous year. The total number of snooping attacks that were detected in the year 2013 stood at close to 45000 and this resulted to financial losses that stood at over 5.9 billion dollars. This meant that there was a 1% increase in the number of snooping attacks from the year 2012 to 2013. The total number of snooping attacks that were observed in the first quarter of the year 2014 stood at 125215 and this was a 10.7% rise from the fourth quarter of 2013. Over 55% of the snooping sites have a similar name to that of the target website in order to dupe the user. Research has shown that payment systems and services are mostly targeted by the snooping attackers in the financial industry [10].

3. THE MECHANISM OF SNOOPING ATTACKS

The snooping mechanism is shown in Figure 1. The fake site is the clone of the genuine website that the hackers target. It always has input fields such as the text area where the targeted user enters his or her personal information which is then transferred to the hacker [11]. The hacker then steals this personal information from the unknowing user through the following steps:

Developing the snooping website; this is the initial step that the hacker takes by identifying the target or the organization. The attacker then gathers comprehensive information about the company by vising the website of the organization. The attacker then uses the information to develop a similar website [12].

Sending the URL. Here the hacker, creates an email that is bogus and sends to many users. In the email, the hacker has attached the URL of the fake website. The attacker can also spread the link of the snooping site using blogs or social media sites to reach many users [13].

Stealing the confidential information: when the unknowing user clicks on the embedded link, the fake website opens in the browser. The fake site has a fake login interface or login form that the attacker uses to steal personal information from the victim. Moreover, the attacker is able gain access to confidential information that the user has filled up [14].

Identity theft: the hacker then uses the personal information obtained for malicious purposes. For instance, the hacker may make purchases online using the credit card information of the unknowing victim [15].

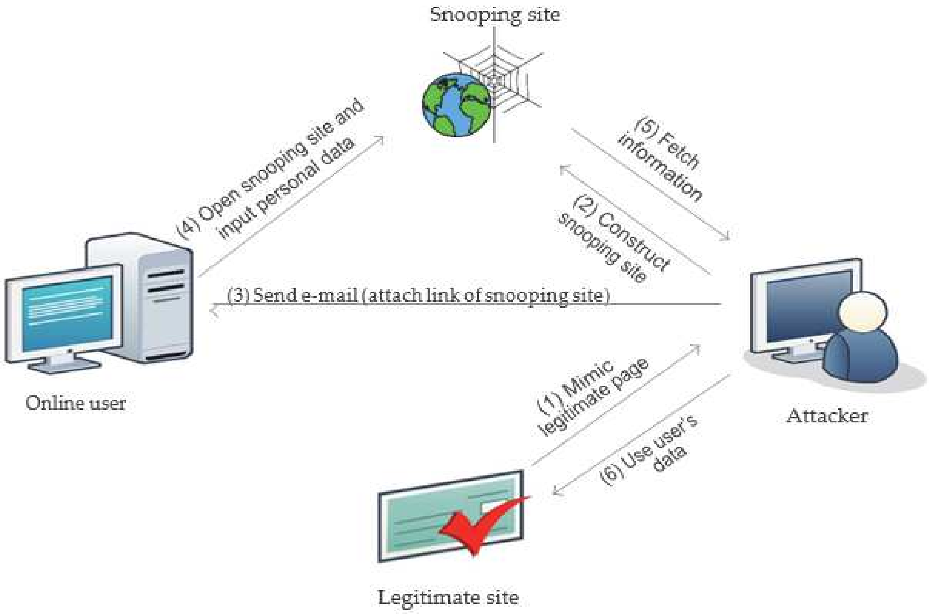
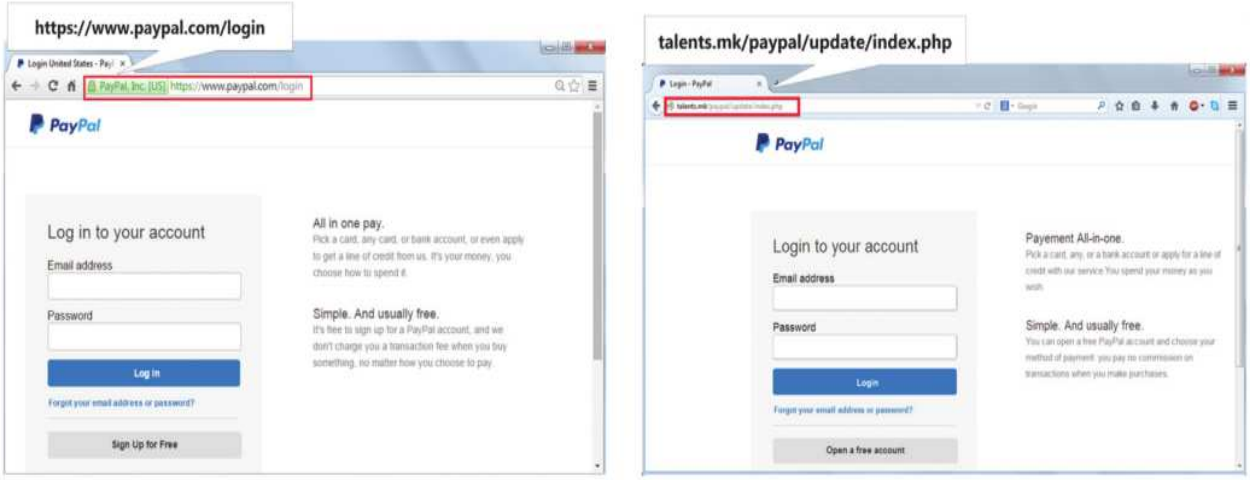

Snooping site
(4) Open snooping site and input personal data
(5) Fetch information
(2) Construct snooping site
(3) Send e-mail (attach link of snooping site)
Online user
Attacker
(1) Mimic legitimate page
(6) Use user's data
Legitimate site

Figure 1: The snooping mechanism

3.1 The Taxonomy of Visual Similarity Snooping Attacks

The attacker carries out the snooping attack by using social engineering mechanisms and technical subterfuge. With the social engineering mechanisms, the hackers manage to attack the unknowing users by sending out bogus emails to thousands of unsuspecting users. The attackers normally convince the recipients of the emails to respond to the emails by keying in their names, their bank details, credit card firms and e-retailers [15]. The technical subterfuge mechanism installs malware into the computer system of the user and, in the process, personal and confidential information is stolen by the use of key logger spyware and Trojan malware. The malware also misdirects the users to websites that are fake or proxy servers [16]. The hackers embed malicious links or fraudulent links/ URLs in the emails which install malicious applications or software in the system of the user. The malicious software then collects confidential data from the system and sends it back to the hackers. The hackers can also remotely access the computer system of the user and then gather the data that they deem necessary [15].

A person can easily become a snooping attack victim due to the high visual resemblance of the visual similarity snooping site with the genuine site because of the page set up, image layouts, font color and size and the content. Figure 2 in an example of a fake and a genuine page of PayPal. The websites have the same visual appearance, however, on a keen a look, one can observe that the URLs are different. People are not always careful to take note of the URL and the SSL Certificates of the sites [17, 18].


https://www.paypal.com/login
Login United States - PayPal
PayPal, Inc. [US] https://www.paypal.com/login
PayPal
Log in
Sign Up for Free
talents.mk/paypal/update/index.php
Login - PayPal
talents.mk/paypal/update/index.php
Google
PayPal
Login
Open a free account

Log in to your account

Email address

Password

Forgot your email address or password?

All in one pay.

Pick a card, any card, or bank account, or even apply to get a line of credit from us. It's your money, you choose how to spend it.

Simple. And usually free.

It's free to sign up for a PayPal account, and we don't charge you a transaction fee when you buy something, no matter how you choose to pay.

(a)

Login to your account

Email address

Password

Forgot your email address or password?

Payement All-in-one.

Pick a card, any, or a bank account or apply for a line of credit with our service You spend your money as you wish.

Simple. And usually free.

You can open a free PayPal account and choose your method of payment. you pay no commission on transactions when you make purchases.

(b)

Figure 2: a. Genuine PayPal webpage and b. Snooping webpage of PayPal

If the hacker does not manage to copy the visual resemblance of the website being targeted, then the probability of the users inputting their credentials is very small [19]. The aim of the hacker is to fool the users using the following ways:

i. Through visual appearance; the snooping website has a similar look to that of the authentic website. The hackers steal a copy of the source code used to build the legitimate website to develop the fake website.
ii. Address Bar; the hackers also hide the URL or the address bar of the site using an image or a script. This makes the users think that they are keying information on the legit site.
iii. Embedded objects; the hackers also utilize objects that are embedded, for example, scripts and images to conceal the HTML code or the textual content from the snooping detection mechanisms.
iv. Favicon similarity; this refers to an image that is linked to a specific site. A hacker can copy the image of the website that is targeted. If the shown favicon in the address differs from the current website, then it is regarded as a snooping attempt.

Research conducted by Dhamija et al., [20] on different users to identity if a website is genuine or a snooping site established that 90% of the users were not able recognize snooping attacks. A majority of the users judged wrongly the website using its visual appearance. The scholars also found out that even the experienced users could easily be duped through the visual similarity of the illegitimate website. They also noted that 23% of the participants do not take time to view the site's address bar. Thus, it can be concluded that if the appearance of the snooping website looks exactly like that of a legit website and with a different domain, then the users can easily be fooled by the snooping attackers [21].

4. MECHANISMS FOR DETECTING SNOOPING ATTACKS

The following are mechanisms that have been devised for the purpose of detecting snooping attacks:

i. Attribute based
ii. Identity based
iii. Content based
iv. Character based

i. Attribute based anti-snooping Technique

The attributes based anti-snooping technique executes every proactive and-anti-snooping technique defenses. This method has also been reinforced in Phish Bouncer Tool [22]. A comparison of images visiting the website will be done using the image attribution check and also check for sites that are already registered under the Phish chucker-out. The HTML cross-link checks for responses that originate from websites that are not registered and counts the different links that are not from registered websites [23]. When there is a big number of cross-links, it shows there is a snooping site. In the feeder check of false information, the false data is keyed in and if this information is accepted by the website, then there is a high chance the link is also a snooped link. The anti-snooping suspicious check analyzes and validates the certificates that are given throughout the SSL handclasp and this carries on to the day usage by logging in for certification authority as time goes by [23].

Pros: this technique takes into consideration a lot of checks so that it can identify snooping websites in comparison to the other methods. The method can also detect snooping attacks that are known and those that are not known [24].

Cons: Because the technique carries out a lot of checks for authentication of a website, there is a high probability of slow response time. [23]

ii. Identity based anti-snooping techniques

This mechanism makes use of the methodology of mutual authentication where an online entity and every user confirms each other's identity through test suggestibility or handclasp. The methodis associated with the technique of nursing ant snooping which makes use of partial credential sharing alongside shopper filtering mechanism to hinder the attackers from pretending to be legit online users [25]. Mutual authentication is followed in this method hence there is no need for the users to re-enter their details. Therefore, the method of using passwords has never changed between the users as well as online entities with the exception of the first method of setting it up [23].

Pros: this technique provides for mutual authentication for the client and server side. Making use of this technique does not expose the personal details of a user, for example, the password that is set up except for the initial time that it is set up [26].

Cons: Using this technique, if the attacker is able to access the user's computer and then disable the browser plugins, it ends up being compromised [23].

iii. Content based anti-snooping approaches

The Gold Phish tool executes this technique and then utilizes google as the program of the computer. This technique then offers seniority to firm websites on the internet. It has been confirmed that snooping web-content for a small amount of time can obtain low ranks in terms of internet search and this then becomes the foundation for this technique [27]. The approach of planning can be reduced to three main steps. The major step is capturing an image of the website in the user's application. The step that then follows is using the optical character mechanisms for converting the image captured to text that is machine readable. The third step entails inputting the text that is reborn into a research engine in order to obtain results and analyze the rank of the page [28].

Pros: Overall, the Gold Phish does not lead to false positive. Also, it offers zero day snooping.

Cons: The Gold Phish technique slows down the process of rendering a webpage. In addition, it is also vulnerable to attacks on the Google's Page Rank algorithm and the search service [29].

iv. Character based anti-snooping approaches

Many times when hackers attempt to steal data from users, they do so by enticing the users to click on URI and hyperlinks that they have embedded in snooped emails. A hyper link is made up of the format; <ahref='URI'. Anchor text, \a. whereby the URI (Universal Resource Identifiers) offer the real link to where the user shall be guided to. The anchor text refers to the text that shall be displayed in the web browser and stands for the visual link [30]. This approach makes use of hyperlink characteristics in detecting the links that are snooped. Link Guard refers to a tool that implements and executes this methodology. After many snooping websites are analyzed, the hyperlinks are then grouped into different categories. In order to detect snooping websites, the Link Guard tool initially obtains the DNS names, and if the names are not the same then it is a snooping attack [31].

The weakness of this methodology is that it can lead to false positives because it uses decimal IP addresses that are dotted in place of domain names. Nonetheless, this may be appropriate in some special situations [32].

5. CONCLUSION AND FUTURE WORK

With advancement in technology, the recent years have come with a drastic increase in the sophistication and the number of email snooping attacks. Several techniques have been developed to detect and prevent snooping attacks such as attribute based anti-snooping techniques, identity based anti-snooping techniques, content based anti-snooping approaches, and character based anti-snooping approaches. Visual similarity- based snooping involves sending of large amounts of emails that are spoofed asking the targeted users to click the links embedded in emails. By just a mere glance, the hyperlinks in the emails are normally difficult to suspect and this makes it easy for the victim to click on them without their knowledge. Future work on the visual similarity snooping technique should entail creating improved ways that will detect the malicious links and attachments in the snooped emails and have the capability of deleting them. Another improvement that should be done on the technique is applying techniques of machine learning to make the visual similarity snooping technique adaptive in nature.

Bibliography

[1] G. A. M. a. S. Yarmohammadi, "Detection of phishing attacks in Iranian e-banking using a fuzzy-rough hybrid sys," *Applied Soft Computing,* vol. 35, no. 1, pp. 482-492, 2015.

[2] A. Kak, Mounting Targeted Attacks for Cyber Espionage with Trojans and Social Engineering, Purdue University, 2020.

[3] A. A.Tewari, "“Recent survey of various defence mechanisms against phishing attacks," *Journal of Information Privacy and Security ,* vol. 12, no. 2, pp. 3-13, 2016.

[4] J. M. E. J. V.-F. a. G. F.-E. Pavía, "Credit card incidents and control systems," *International Journal of Information Management ,* vol. 32, no. 6, pp. 501-503, 2012.

[5] S. Z. Nur, S. Deris , F. A. R. Mohd Faizal , F. Ahmad , I. S. W. D. Wan , K. Shahreen and S. Tole , "Phishing detection system using machine learning classifiers,", vol. 17, Indonesian Journal of Electrical Engineering and Computer Science, 2020, p. 1165~1171 .

[6] G. A. a. S. A. Montazer, "Detection of phishing attacks in Iranian e-banking using a fuzzy–rough hybrid system," *Applied Soft Computing,* vol. 1, no. 35, pp. 482-492, 2015.

[7] I. a. C. I. Drigă, "E-banking services–features, challenges and benefits," *Annals of the University of Petroşani. Economics,* vol. 14, pp. 49-58, 2014.

[8] I. W. Stats, "Internet World Stats," Internet World Stats, 2019. [Online]. Available: https://www.internetworldstats.com/stats.htm. [Accessed 24 October 2019].

[9] C. Crane, "20 Phishing Statistics to Keep You from Getting Hooked in 2019 - Hashed Out by The SSL Store," .thesslstore.com, 2019. [Online]. Available: https://www.thesslstore.com/blog/20-phishing-statistics-to-keep-you-from-getting-hooked-in-2019/. [Accessed 24 October 2019].

[10] W. O.-R. J. C. K. a. J. S. Kim, "The dark side of the Internet: Attacks, costs and responses," *Information systems,* vol. 3, no. 36, pp. 675-705, 2011.

[11] F. A. M. a. A. B. Alkhateeb, "Bank web sites phishing detection and notification system based on semantic web technologies," *International Journal of Security & Its Applications ,* vol. 6, no. 14, pp. 1-14, 2012.

[12] S. A. &. R. Greenstadt, "PhishZoo: Detecting Phishing Websites By Looking at Them," *Detecting Phishing Websites By Looking at Them,* vol. 2, no. 6, pp. 1-8, 2015.

[13] M. M. &. P. K. A. Gaurav Varshney, "A survey and classification of web phishing detection," *SECURITY AND COMMUNICATION NETWORKS,* vol. 1, no. 9, pp. 6266-6284, 2016.

[14] A. Y. &. Y. M. Masanori Hara, "Visual Similarity-based Phishing Detection without Victim Site Information," vol. 1, no. 3, pp. 1-7, 2015.

[15] A. K. a. B. B. G. Jain, "Phishing detection: analysis of visual similarity based approaches," *Security and Communication Networks,* pp. 1-21, 2017.

[16] Y. Zhou, Y. Zhang, J. Xiao and Y. Wang, "Visual Similarity Based Anti-phishing with the Combination of Local and Global Features," *In 2014 IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications,* vol. 1, no. 3, pp. 189-196, 2014.

[17] C. Yi-Lun, L. Iuon-Chang and C. Hung-Chieh, "The features of phishing detection based on judgment user device,", Brunei, Bandar Seri Begawan: 4th International Conference on Computer Technology and Science, 2015.

[18] Y. Peng, Z. Guangzhen and Z. Peng, "Phishing Website Detection based on Multidimensional Features driven by Deep Learning,", IEEE, 2019.

[19] L. Cheng-Chi, . L. Chia-Hsin and H. Min-Shiang , "Guessing Attacks on Strong-Password Authentication Protocol,", vol. 15, International Journal of Network Security, 2013, pp. 64-67.

[20] R. J. D. T. a. M. H. Dhamija, "Why phishing works," *Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM,* vol. 5, no. 3, pp. 581-590, 2016.

[21] N. A. G. S. L. a. K. B. Arachchilage, "Phishing threat avoidance behaviour: An empirical investigation," *Computers in Human Behavior ,* vol. 60, no. 1, pp. 185-197, 2016.

[22] E. Medvet, E. Kirda and C. Krügel, "Visual-Similarity-Based Phishing Detection," eurecom.fr, 06 July 2018. [Online]. Available: http://www.eurecom.fr/en/publication/2515/detail/visual-similarity-based-phishing-detection. [Accessed 12 12 2019].

[23] M. S. K. P. a. U. S. Deshmukh, "Different Techniques for Detection of Phishing Attack.," *International Journal of Engineering Science and Computing,* vol. 7, no. 4, pp. 1-4, 2017.

[24] G. S. &. Kuppusamy, "PhiDMA – A phishing detection model with multi-filter approach," *Journal Of King Saud University-Computer and Information Sciences ,* vol. 32, no. 1, pp. 99-112, 2020.

[25] B. Jongman, "Resources Recent Online Resources for the Analysis of Terrorism and Related Subjects," *Perspectives on Terrorism,* vol. 13, no. 2, pp. 156-189, 2019.

[26] R. W. &. W. J. K. Ausen, "Extrusion die element, extrusion die and method for making multiple stripe extrudate," *U.S. Patent and Trademark Office,* vol. 9, no. 2, pp. 327-429, 2016.

[27] C. M. L. J. &. A. D. Tian, "Phishing Susceptibility across Industries: The Differential Impact of Influence Techniques.," *Proceedings of the 13th Pre-ICIS Workshop on Information Security and Privacy,* vol. 1, no. 1, pp. 1-20, 2018.

[28] A.V.R.Mayuri, "Phishing Detection based on Visual-Similarity," *International Journal of Scientific and Engineering Research,* vol. 3, no. 6, pp. 1-5, 2012.

[29] J. R. &. W. P. M. Hollenbeck, "Harking, sharking, and tharking:," *Making the case for post hoc analysis of scientific data.,* vol. 5, no. 2, pp. 5-18, 2017.

[30] A. S. A. &. D. W. B. Wright, "The big phish: cyberattacks against US healthcare systems," vol. 3, pp. 1115-1118, 2016.

[31] R. &. M. B. Butler, "Assessing the information quality of phishing-related content on financial institutions' websites," *Information & Computer Security,* 2018.

[32] M. Eoyang, "Beyond Privacy and Security," *The Role of the Telecommunications Industry in Electronic Surveillance,* vol. 9, no. 3, p. 259, 2017.

[33] P. a. S. M. Pasricha, "Electronic crime in Indian banking," *Sai Om Journal of Commerce and Management ,* vol. 1, no. 11, pp. 7-14, 2014.