International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-9 Issue-1, May 2020

Detection of Phishing Attacks using Content Analysis in the Cloud

Sweta Mittal, Jayasimha S R

Abstract: ***In the world of internet and technology, technical advancement is widely accepted by both types of users - with and without technical knowledge. Advancement in technologies also brings in different risks involved along with it. These risks involve risks of being compromised at any point of time, leading to identity theft or financial loss or loss of very confidential information. Phishing attacks are one such kind of attack which can trap anyone into it, let it be a novice user or a sophisticated user. This paper involves what phishing attacks are, how the phishers target cloud services, how they deceive users, how the phishers send phish sites to its target. It also includes the background process that happens in normal scenarios and during phishing, a proposed mechanism which can be used for detection, safety measures which if taken can reduce the chances of falling in the trap and mechanisms used by researchers in order to detect and prevent phishing sites.***

Keywords - phishing, mobile cloud computing, content analysis, machine learning algorithms

I.INTRODUCTION

Have you ever done shopping and after returning home, got a mail or a message stating you have got a cashback or a foreign trip? Have you ever received messages stating that a xyz company is offering a limited period deal and a link where you can avail the link. If yes, beware of such messages as those can be sent by phishers performing a phishing attack. Phishing is a technique of acquiring sensitive information from people through social engineering tricks. In phishing, the phisher tries to exploit human vulnerability instead of targeting software vulnerability [1]. The phishers send a counterfeit email to the users which appears to be a legitimate email. This email contains a link to a phishing site which resembles a legitimate website and asks the users to enter sensitive data. These can be username, password, credit card number, etc., which are within the privacy boundaries of the user.The users are tricked to enter their personal information and unknowingly fall in the trap set by the phisher. Phishing has become one of the internet security threats globally. Anyone can fall in the trap of the phishers, even the most sophisticated users.

Manuscript received on April 02, 2020.
Revised Manuscript received on April 20, 2020.
Manuscript published on May 30, 2020.

***** Correspondence Author

Sweta Mittal*, PG Student Department of Master of Computer Applications RV College of Engineering, Bangalore

Jayasimha S R, Assistant Professor Department of Master of Computer Applications RV College of Engineering, Bangalore

According to the statistics presented by the Anti-Phishing Working Group(APWG), the frequency of phishing attacks keeps increasing [2], also reports from Kaspersky Lab states that,in comparison from phishing attacks identified in 2016, it has increased by 47.48% [3]. Phishing attack is performed to achieve a goal set by the phisher. As indicated by the past few years of phishing trends, the major target of phishing is financial institutions [4, 5]. Most of the phishers aim at financial gain; they disguise themselves as financial institutions and attack its customers or target the employees of the financial institutions. These phishing scams result in both financial as well as economic loss for the victims [6].

II.CLOUD SERVICES AS A TARGET OF PHISHERS

In the modern era, almost everything has moved to online. Also the introduction of cloud computing has changed the way how people and organizations work. Various companies store their data in the cloud and use various services given by the cloud service providers like launching instances, hosting servers, etc. Also meetings in many companies occur online over a meeting application. This increases the frequency of the people's interaction with the technology. Since many organizations do their computing in the cloud , the hackers try to target the organization's cloud infrastructure. For that, the easiest way is to trick an employee and gain his/her credentials and then impersonify themselves as that employee and gain access. For this the hacker does research on employees and figures out which employees would fall in the trap. Then the hacker sends them phishing emails and retrieves their login credentials and uses them to perform some advanced attacks against the organization [7,8]. Many advanced attacks such as ransomware, Advanced Permanent Threats(APT), etc. begin with phishing. The phishers try manipulating the users of cloud-provided services by disguising themselves as cloud service providers, for example, by sending fake usage bills to users. Most cloud services provided to the users still use a simple authentication process that involves username and password. Some institutions like financial institutions use a secondary authentication process to validate the users and prevent it from attacks such as phishing [9]. Mobile cloud computing is a new approach taken where the development of mobile computing is combined with the cloud-based computing and services and network [10]. In this approach, all the computing, storage and applications are delivered through cloud. With the increase in mobile cloud computing, more and more devices are dependent on cloud for computing and storage. In such computing, authentication becomes a critical security issue. The user can be verified with the mobile device.

In the current scenario, in the approach used for authentication, the hash value or the encrypted value of the password is sent to the user's device [11,12]. This process can lead to the intruders to capture the password. Also this can motivate the phishers to design fake emails or websites and send it to users to capture the credentials.

III.HOW IT DECEIVES USERS

Many people don't have knowledge about the computer science field. They don't have knowledge about domain names, how they are used to address websites, how they are structured, how the legitimate urls look like, etc. This makes them easy to fall in the trap as they are unable to figure out the phishing sites and trust them.

Many people who are aware of phishing attacks and have technical knowledge still sometimes fall in the trap laid down by the phishers. It's all because of the skills of the phisher in presenting it to the victim. The main goal of the phisher is to fool the victim, by imitating valid images, text and windows using visual deception tricks.

Best way to deceive is by using an image of a legitimate link and linking the image to a different fraud website [13]. Visual deception using text is done mostly in URLs. The phisher tries to use a URL which looks almost similar to the original URL to host the phish site. For example, a phish website of a bank was hosted at "www.bankofthevvest.com", where the phisher used 2 "v"s in the place of a 'w' [14].

Also the phishers can use social networks like Facebook, Instagram, LinkedIn, etc. and with the help of social engineering, try to do some background study. The victim's personal profile is analysed along with the work history, activities and interests. Using this analysis, the phishers can plan out an approach to attack the victim. Since the approach chosen by the phisher is based on the study, there are less chances of it to fail and the victim falling in trap.

IV.HOW PHISHERS PERFORM PHISHING

Phishers perform the attack using a medium which is used oftenly by the victim. Through these mediums the phisher can send the url of the phish site. Some mediums are listed as below :-

1) Through emails:- It is the most common and known medium used by the hackers. The phish site can be embedded in an email beautifully designed for the user to believe it and click on the link.

2) Through instant messaging applications:- Another common way used by the hackers which is similar to emails.

3) Through standalone applications:- The hacker can design a mobile application or a desktop application and forward it to the victim through some medium. When the victim installs it and uses it, a phishing attack can be performed without letting the user detect it.

4) Through DNS cache poisoning:- If the phisher gains access to the network, they can attack the Domain Name Systems(DNS) of the network and poison it so that all the requests going to a particular IP is redirected to another IP which loads the phish site [4]. In this attack, the phisher only needs to host the website in its system and add the IP address of its system in the DNS and wait for the victim to fall in the trap.

The phishers can take the advantage of the small display of smartphones. The phishers can share the link through a chat application where mostly the chances are high of the victim opening the website in mobile browsers. In such cases, the URL becomes unnoticed and the page is not properly displayed. Also in normal cases, people probably don't install additional extensions in their mobile phone browsers. So phish sites don't get detected by the browsers and hence the victim falls in the trap of phishers.

V.THE LOGIC BEHIND IT

Normal web applications follow the flow shown in Fig 5.1. All the request from the user goes to the DNS/router which does the mapping of domain name and IP address. Then the request is forwarded to the actual web server hosting the site.

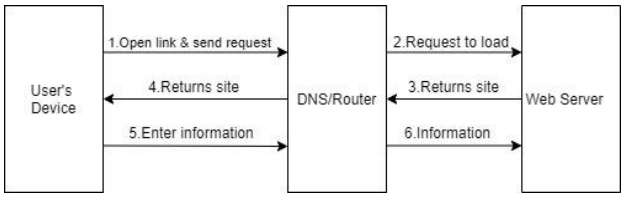

Fig. 5.1 Normal web applications

A phish site works with the mechanism shown in Fig 5.2. When the victim clicks on the link of the phish site which is sent by the phisher, the victim device sends the request to the DNS which maps the phish site url to the phish site IP address. In case of DNS poisoning, the phisher hacks the DNS and maps the url of the legitimate site to the IP address of a phish site web server.This request is then forwarded to the phisher's web server. The phisher's web server returns the phish site back to the user. If after seeing the website, the victim gets convinced it is a legitimate website, then the victim enter's credentials or sensitive information and tries to login. That moment these data are sent back to the hacker's web server and the user is compromised.

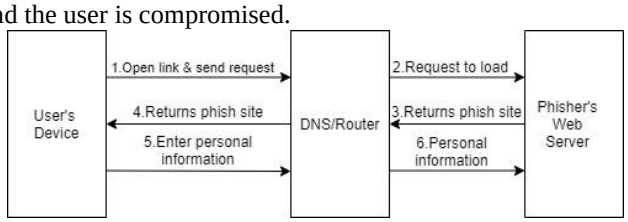

Fig. 5.2 Phish web applications

This chain can be broken if there is a phishing prevention tool introduced in the user's device or between the user's device and the DNS/router. As soon as the user tries to open the link, the tool will first verify the link and check whether it's a phish url or not. If it is a phish url, the request can be immediately blocked and reported to the user.

There are many phishing prevention tools available like cloudphish, netcraft, phishtank, phishprotection, etc.

VI.PROPOSED SYSTEM

In the proposed system as shown in Fig. 6.1, there will be a phishing prevention tool between the user's device and the DNS/Router. Here the phishing prevention tool will work in two phases. In the first phase, when the request is sent from the user's device, the phishing prevention tool will take the URL from the request and check the URL, whether it is a phish URL or not. If it identifies the URL as a phish URL, it will drop the request and notify the user. If the URL is not a phish URL, the tool will forward the request to the DNS/router. This request will reach the web server in which the website is hosted which will forward the website. Before the website reaches the user's device, the phishing prevention tool will analyse the content of the website using a content analysis approach. In this approach, a model will be created which will have unique and already tracked down patterns and signatures which identifies that the website is not a legitimate website and will also be having all the website catalogs that are being identified as phishing websites. It will also include checks like buttons without links, spelling errors, ranking of the website, content of the website, sources of the images that are available in the repository of the legitimate website. This approach identifies and tracks the websites which are turning out to be a phishing website. Those websites are blocked and a notification will be sent to the user's email id.

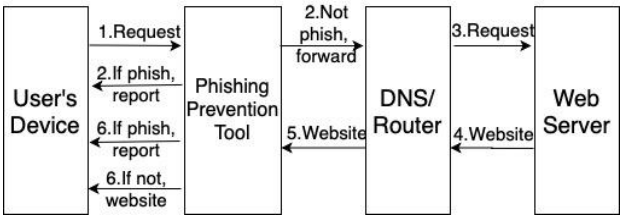

Fig. 6.1 Proposed System

All the necessary components which proves that a websites is legitimate whether it be a request generated from a genuine server, or the URL containing the request which is being generated by the user,from first step till the last step all the crucial components will be checked and will be ensured that no checks should be left in order to comment that website is legitimate or not.

VII.SAFETY MEASURES THAT CAN BE TAKEN

1) Checking the URL of the website:- Before entering sensitive data to any website, first check the URL of the website. It is the best way to detect a phish site. Big organizations such as a bank, social networking site, etc. has it's domain name which no one else can use to host a website. Hackers try to use a similar domain name which looks similar to the original domain name so that the user gets fooled and enters sensitive data.

2) Using phishing prevention extensions in web browsers:- There are many phishing prevention tools available for web browsers which can be downloaded and used.

3) Phishtank:- Whenever there is any URL which the user is not sure about, a website named Phishtank can be used to check whether it is a phish site or not.

4) Cloudphish:- Cloudphish is an extension provided by chrome browser which helps the users detect phish sites. It provides end to end validations to all the emails received through Gmail or Outlook. It protects it's users from all kinds of phishing emails. It does not come preinstalled with the browser. It has to be manually installed by the user.

5) Netcraft:- Netcraft provides an anti-phishing toolbar for Chrome browser, Internet Explorer and Firefox . It has to be manually installed in the browser. Netcraft also has a search tool where the users can use to search for hosting information related to a website [15].

6) Avoid clicking on unnecessary links and don't open any link which seems to be fraud.

7) Before believing an email received from an organization, which can lead to financial loss or identity loss, contact the customer care or anyone from the organization to verify whether the email is valid or not.

8) Before logging in with an email id, first verify the site:- Since many accounts are linked to email accounts, phishers aim at gaining email credentials of the user. If email id is compromised, the accounts in different portals linked to it will also be compromised. In order to do that, phishers design the phish site similar to an email sign in page, which looks like a legitimate site.

VIII. MECHANISMS INTRODUCED TO DETECT THE PHISHING ATTACKS

In this era of data science and machine learning, various machine learning techniques are taken and are tested by the researchers to find solutions which can detect phishing sites. In machine learning, a model is defined using training data and then the model is used for computations. The advantage of prediction power is taken in the machine learning approach. Also there are various classification algorithms in data mining like decision tree, naive Bayes, Bayesian net, etc. that can be used to classify the various types of phish emails and phish sites and then test the approach taken by checking the number of false positives [16]. Phishing problems can be detected using various approaches. Blacklist is an approach in which the phishing url is compared with a list of databases containing phishing urls for a match [17]. Each time a new phishing url is introduced, this mechanism fails to detect the phish site.
Chen et al. designed a mechanism where they detected phishing sites using screenshots of the site [18,19]. For this, Contrast Context Histogram(CCH) was used to define the pictures and find the similarities between 2 websites using euclidean distance.

This mechanism provided 95-99% accuracy. Since the approach is time consuming and will be slow in real time analysis, it cannot be used in real time. There is an Anti Phishing Simulator designed [7], which can be installed in each system.

Any malicious emails arriving at the system integrated email address is avoided by the application. This system also uses the existing database and accesses keywords which can be used to determine the text in emails. Alexa is another way to detect phishing sites. It rates a website on the basis of its popularity. The more the website is visited, the more popular the website is. Generally a phish site will be a newly developed website and will not have much popularity. It can be used to measure the website traffic and used mostly by researchers for detection [20].

IX.CONCLUSION

Researchers work on finding solutions using various machine learning and data mining approaches in order to detect phish sites and save users from being compromised. But the phishers stand one step ahead. They try out ways through which they can break through such mechanisms and succeed in their motive. They also work hard to make emails and phish sites so real that anyone who sees it believes it to be a legitimate email or site. Awareness regarding phishing needs to be spread among the users so that they know what it is and how they can reduce the chances of being compromised by taking safety measures and precautions.The phish prevention tool will work on two phases, that brings in more accuracy. If a phish site is detected in the first phase, then it saves time. If not detected in the first phase, then it will be detected in the second phase.

REFERENCES

1. A. Alswailem, B. Alabdullah, N. Alrumayh and A. Alsedrani, "Detecting Phishing Websites Using Machine Learning," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2019, pp. 1-6
2. AO Kaspersky lab. (2017). The Dangers of Phishing: Help employees avoid the lure of cybercrime. [Online] Available: https://go.kaspersky.com/Dangers-Phishing-Landing-Page-Soc.html [Oct 30, 2017]
3. ”Financial threats in 2016: Every Second Phishing Attack Aims to Steal Your Money” Internet: https://www.kaspersky.com/about/press- releases/2017 financial-threats-in-2016. Feb 22, 2017 [Oct 30, 2017]
4. W. D. Yu, S. Nargundkar and N. Tiruthani, "A phishing vulnerability analysis of web based systems," 2008 IEEE Symposium on Computers and Communications, Marrakech, 2008, pp. 326-331
5. M. Khonji, Y. Iraqi and A. Jones, "Phishing Detection: A Literature Survey," in IEEE Communications Surveys & Tutorials, vol. 15, no. 4,Fourth Quarter 2013, pp. 2091-2121
6. M. Blasi, ”Techniques for detecting zero day phishing websites.” M.A. thesis, Iowa State University, USA, 2009
7. M. Baykara and Z. Z. Gürel, "Detection of phishing attacks," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018, pp. 1-5
8. Ş. Şentürk, E. Yerli and İ. Soğukpınar, "Email phishing detection and prevention by using data mining techniques," 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, 2017, pp. 707-712
9. Ajey Singh, Dr. Maneesh Shrivastava, ”Overview of Attacks on Cloud Computing
10. ,” International Journal of Engineering and Innovative Technology (IJEIT), Volume 1, Issue 4, April 2012
11. Ellappan, Munivel & A, Kannammal. (2019). New Authentication Scheme to Secure against the Phishing Attack in the Mobile Cloud Computing. Security and Communication Networks. 2019,pp. 1-11
12. M. Alizadeh, S. Abolfazli, M. Zamani, S. Baaaharun, and K. Sakurai, “Authentication in mobile cloud computing: a survey,” Journal of Network and Computer Applications, vol. 61, 2016 , pp. 59– 80
13. B. K. Chaurasia, A. Shahi, and S. Verma, “Authentication in cloud computing environment using two factor authentication,” in Proceedings of the Third International Conference on Soft Computing for Problem Solving, vol. 259 of Advances in Intelli- gent Systems and Computing, Springer, New Delhi, India, 2014, pp. 779–785
14. J. Chen and C. Guo, "Online Detection and Prevention of Phishing Attacks," 2006 First International Conference on Communications and Networking in China, Beijing, 2006, pp. 1-7
15. Dhamija, Rachna & Tygar, J. & Hearst, Marti. (2006), “Why phishing works”, Conference on Human Factors in Computing Systems - Proceedings. 1, pp. 581-590
16. Wu, Min & Miller, Robert & Garfinkel, Simson. (2006), Do security toolbars actually prevent phishing attacks?, Proceedings of the SIGCHI conference on Human Factors in computing systems CHI. 06., pp. 601-610.
17. Arun Kulkarni and Leonard L. Brown III, “Phishing Websites Detection using Machine Learning” International Journal of Advanced Computer Science and Applications(IJACSA), 10(7), 2019
18. A. K. Mishra, A. K. Tripathy and S. Swain, "Analysis and Prevention of Phishing Attacks in Cyber Space," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 430-434
19. S. Afroz and R. Greenstadt, "PhishZoo: Detecting Phishing Websites by Looking at Them," 2011 IEEE Fifth International Conference on Semantic Computing, Palo Alto, CA, 2011, pp. 368-375
20. K. Chen, J. Chen, C. Huang and C. Chen, "Fighting Phishing with Discriminative Keypoint Features," in IEEE Internet Computing, vol. 13, no. 3, May-June 2009, pp. 56-63
21. L.A.T.Nguyen, B.L.To, H.K.Nguyen, M.H.Nguyen, ”Detecting phishing web sites: A heuristic URL-based approach,” in 2013 Inter- national Conference on Advanced Technologies for Communications (ATC 2013), 2013, pp. 597-602

AUTHORS PROFILE

Sweta Mittal, a final semester student in the Department of MCA at RV College of Engineering®, Bangalore. Currently she is doing her internship in Bidgely Inc. as a data engineer. Her area of interest is cyber security and data science. She is more familiar with python and java as compared to other programming languages.

Prof. Jayasimha S R, working as an assistant Professor in the department of MCA at RV College of Engineering®, Bangalore. He served in the institution from 2013 to till the date. Currently he submitted his PhD Thesis to the VTU. His area of interest is cloud computing. He published more than 20 papers in national, international conferences and international Scopus indexed journals.