Digital Watering Hole Attack Detection Using Sequential Pattern

T. Subburaj[1] and K. Suthendran[2]

[1] *Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil - 626126, Tamilnadu, India*
[2] *Department of Information Technology, Kalasalingam Academy of Research and Education, Krishnankoil - 626126, Tamilnadu, India*
E-mail: shubhurajo@gmail.com; k.suthendran@klu.ac.in

Received 23 February 2018; Accepted 12 March 2018;
Publication 20 March 2018

Abstract

Internet plays a vital role in day to day communication, business transactions etc and thus unavoidable. But many of the users are lagging in using the same in a secured manner which increases the possibility of attack. In 2017, Attackers had targeted dozens of global banks with new malware. Watering hole attacks attempt to infect more than 100 organizations in 31 different countries. It is becoming very difficult to detect and prevent the cyber attacks, since new attacks are increasing day by day. A watering hole attack is a computer attack in which the attacker aims to victim the kind of websites that the target group go to often and checks these websites for vulnerabilities. After that by injecting Java Script or HTML redirects the victim to a separate site hosting the exploit code for the chosen vulnerability. In this paper, a novel method is proposed to detect the possibility of watering hole attack using support count and confidence of the sequence pattern mining. Further by analysing the website URL links, alarming the users about the watering hole attack.

Keywords: Watering hole, Relative confidence, Cyber attack, Sequence pattern, Phishing.

Journal of Cyber Security, Vol. 7_1, 1–12.
doi: 10.13052/jcsm2245-1439.711

1 Introduction

In today's situation, internet is not only used for the web mail and chat but also extended to the field of studies, media, business and many more. It makes people life easier, but they are not much familiar about the security. When the internet usage increases, the possibility of attacks is also increasing. Since the attackers are developing themselves by gaining adequate knowledge to meet the current trends and also creating the new types of attack. During the past, the attackers were simply used "ILOVEYOU" spam emails to attack the systems. Years passing by they started developing difficult and critically identified programs to attack the targets.

In networking the users are facing lot of issues which are being created by the attackers. The attacker steals the user information's through the network and causes our system as unsecured. Nowadays instead of attacker getting into our network and system they wait for us to come to them. So what actually they are looking for is to find out that where we go, when we work in the office or when we access the other websites outside of private internal network this is called digital watering hole. This is the moment when people go from inside of your network to popular websites people likes to visit for shopping etc. This requires bit of research to determine what sites people in your organisation usually go to. They can make us to come to the watering hole might be able to take the advantage other.

The digital watering hole attack is a more complex form of a Phishing attack. In this attack, attackers wait for the users to come to them.

In 2009, big Multinational Companies viz., Adobe, Google and other manufacturing company were suffered by watering hole attack called "Aurora" [15]. Similarly GhOstNet attack was performed on government agencies and embassies worldwide.

During 2012, the visitors of Council on Foreign Relations website was also targeted using digital watering hole attack that is Zero day Exploit. Some other examples are as following: http://cartercenter.org, http://princegeorgescountymd.gov, http://rocklandtrust.com (Massachusetts Bank), http://ndi.org (National Democratic Institute), http://www.rferl.org (Radio Free Europe/Radio Liberty) [11]. In between June 2012 and July 2012, the RSA sites were hacked by some attackers, and also redirected this website to torontocurling.com

During early 2013, a website of labour department, United States was identified as a targeted website by attackers with an aim to collect labour information. Later, the watering hole attack was found in Rapids7 website. Further, the eye-watch.in also affected by watering-hole attack.

On Nov 2016, National Banking and Stock Commission of Mexico were attacked by some hackers using the watering hole attack, this web site forward to the http://www.cnbv.gob.mx/Prensa/Paginas/Sanciones.aspx.

In late 2016, a Polish bank found computers of an institution with malware. However, there are no reports about any money loss [16].

On February 13, 2017 According to security experts from Symantec and BAE Systems, the recently discovered watering hole attacks aimed at Poland banks are linked to the Lazarus Group. During June 2017, ExPetr attack made comprise of Ukrainian government website [17].

2 Literature Survey

Engin et al. [1] proposed a new method to protect user systems form the website based phishing attack by using new browser extension named Anti-Phish. It tracks user sensitive information and throws warning message whenever he/she tries to give away this information to a distrust website. In [2], author proposed an idea to detect phishing websites. This is the modified version of machine learning application on feature set developed to indicate user targeted deception. Approximately, 860 phishing mails and 6,950 non-phishing mails are traced and detected accurately over 95% of them along with the categorization on the basis of 0.09% of the genuine emails. An effective approach based on fuzzy systems and neural networks [3] was proposed to detect the Phishing attack; this approach achieves 99.6% accuracy compared with existing approaches to identify the attacks. For training and testing the proposed model, they have used nearly 300 data from different datasets of six sources. The authors in [4] have proposed a method for phishing attack detection using prior based learning method. Here they have utilized logistic regression based machine learning classifier for this purpose. Using this approach they achieved 97% of accuracy by demonstrating this work in the anti phishing scenario.

The work in [5] represented as "PhishZoo" a phishing detection approach. It identifies phishing website just by comparing its appearance with trusted website. This approach is also achieves the accuracy in the range 96% similar to other blacklisting approaches. It makes use of computer vision techniques for this purpose. Further this approach also identifies the zero-day phishing attack. Deceptive Phishing is one of the issues in Instant Messengers. Applying this technique one can steal others important information viz., including personal information etc,. Here the user identification is often unknown. Ali et al [6] proposed method for detecting the deceptive phishing web sites viz., Anti phishing detector (APD). Another method proposed for phishing detection is using Hybrid feature selection by Isredza et al. [7]. In this work they have extracted all features using Mbox2xml as a disassembly tool. This hybrid feature selection method has given 94% accuracy.

The authors in [8] have suggested SSL/TLS authentication schemes to safeguard Internet banking customers from phishing attacks based on the SAPIM (Server Authentication using Personal Identification Message). They have X.509 certificate for this purpose initially. In this method all user access the web site based on some filtering condition. Another algorithm named Link guard algorithm to avoid the phishing attacks. It is used to detect the attacks form the hyperlinks. Link guard algorithm analyzes the difference between the visual link and actual link. This algorithm not only detects the phishing attack it also protect the user form the malicious web page. Statistical approach based attack detection is proposed by authors in [9]. In this approach based on the variations in the data flows the attack was detected. Statistical calculations done by the Median calculation to detect the attack is projected by the researchers in [10] instead of Mean.

The above literature witnesses that still many researchers are working towards preventing phishing attacks. However, bad guys keep on developing new methods. Digital Watering hole is one of the phishing attacks. In this work, we propose a novel idea to detect the watering hole attack based on sequential pattern. The remaining paper is organized as follows: Section 3 describes about digital watering hole attack with suitable example; Section 4 briefs about the detection of attack; Section 5 deals with checking the URL links of website; and Section 6 concludes this work.

3 Watering Hole Attack

Attacker identifies the frequently used web site of the target users and infecting those frequently used websites, by adding some HTML or Java Script. The attackers are being in remote infecting the victim systems with Trojan. The watering hole attack is simply successful because the legitimate websites are easily compromised by the attackers and are typically exploiting the zero-day vulnerability for which there are no antivirus or IDS signatures available.

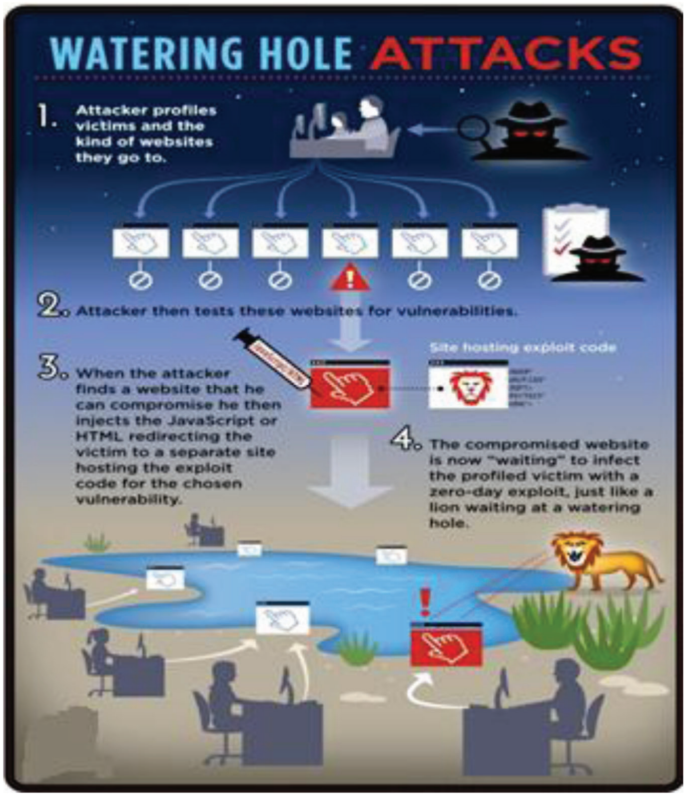

WATERING HOLE ATTACKS
1. Attacker profiles victims and the kind of websites they go to.
2. Attacker then tests these websites for vulnerabilities.
3. When the attacker finds a website that he can compromise he then injects the JavaScript or HTML redirecting the victim to a separate site hosting the exploit code for the chosen vulnerability.
Site hosting exploit code
4. The compromised website is now "waiting" to infect the profiled victim with a zero-day exploit, just like a lion waiting at a watering hole.

Figure 1 Watering hole attack procedure [12].

How the Watering hole attack works?

Watering hole attacks are mainly targets the business peoples, financial institutions, non government organizations and colleges.

The watering hole attack works as follows, it shown in Figure 1.

Step 1: Attacker identifies the victim's system details and also guesses the frequently used websites of victim's.

Step 2: Attacker checks website for the vulnerabilities.

Step 3: Then compromise the website by injecting the JavaScript or HTML and redirect victims to the separate site.

Step 4: The compromised web sites are waiting for the victim user, just like a lion waiting at a watering hole.

4 Detection of Watering Hole Attack

A ‘Sequential Pattern mining’ is used to identify the website which is mostly accessed by target group. The quality of a sequential pattern is measured by support and confidence. Detecting the phishing website is by finding low support and high confidence sequential patterns [12].

Following experiment is used to detect the watering hole attack: Table 1 shows the sequential pattern example.

Assume that Minimum support value as 30%, and confidence value as 70%. Minimum support count 30/100 * 6 = 2. Calculate the association rule of confidence and the relative confidence based on the URL links.

To check the Confidence and relative confidence using the formula 1 & 2:

$$Confidence\,\{X, Y\} \Rightarrow Z = \frac{support\,(< X, Y, Z >)}{support\,(< X, Y >)} \tag{1}$$

$$Relative\ Confidence\,\{X, Y\} \Rightarrow Z = (P\,(X \cap Y) - P\,(X) * P\,(Y))/ \\ (P\,(X)) - (P\,(X) * P\,(Y)) \tag{2}$$

Let us assume that www.kalasalingam.ac.in = a, www.gmail.com = b, www.rajtamil.com = c, www.oneindia.com = d, www.incods2017.in = e. Table 2 show the sequence pattern calculations.

High confidence values are identified from the Table 2. In total, there are four values with high confidence and high relative confidence based on the sequence pattern calculation. High confidence value patterns are identified as attacks domains.

Table 3 shows the list of URLs with high confidence and support values. Based on the sequence pattern matching method the following $\{$***a, b***$\} \Rightarrow$ ***e,***

Table 1 Example of the sequence patterns

User	URL
1	www.kalasalingam.ac.in, www.gmail.com, www.oneindia.com, www.incods2017.in
2	www.gmail.com, www.rajtamil.com, www.incods2017.in
3	www.kalasalingam.ac.in, www.gmail.com, www.oneindia.com, www.incods2017.in
4	www.kalasalingam.ac.in, www.gmail.com, www.rajtamil.com, www.incods2017.in
5	www.kalasalingam.ac.in, www.gmail.com, www.rajtamil.com, www.oneindia.com, www.incods2017.in
6	www.gmail.com, www.rajtamil.com, www.oneindia.com

Table 2 Sequence pattern calculation

URL	Support Count	Confidence	Relative Confidence
$\{a, b\} \Rightarrow c$	2	2/4 = 0.5	2/(4-(4*4) = 0.166
$\{a, b\} \Rightarrow d$	3	3/4 = 0.75	3/(4-(4*4) = 0.25
$\{a, b\} \Rightarrow e$	4	4/4 = 1	4/(4-(4*5) = 0.25
$\{a, c\} \Rightarrow e$	2	2/2 = 1	2/(2-(2*5) = 0.25
$\{a, d\} \Rightarrow e$	3	3/3 = 1	3/(3-(3*5) = 0.25
$\{b, c\} \Rightarrow d$	2	2/4 = 0.5	2/(4-(4*4) = 0.166
$\{b, c\} \Rightarrow e$	2	2/4 = 0.5	2/(4-(4*5) = 0.125
$\{b, d\} \Rightarrow e$	3	3/4 = 0.75	3/(4-(4*5) = 0.18

Table 3 Attacked Sequence pattern

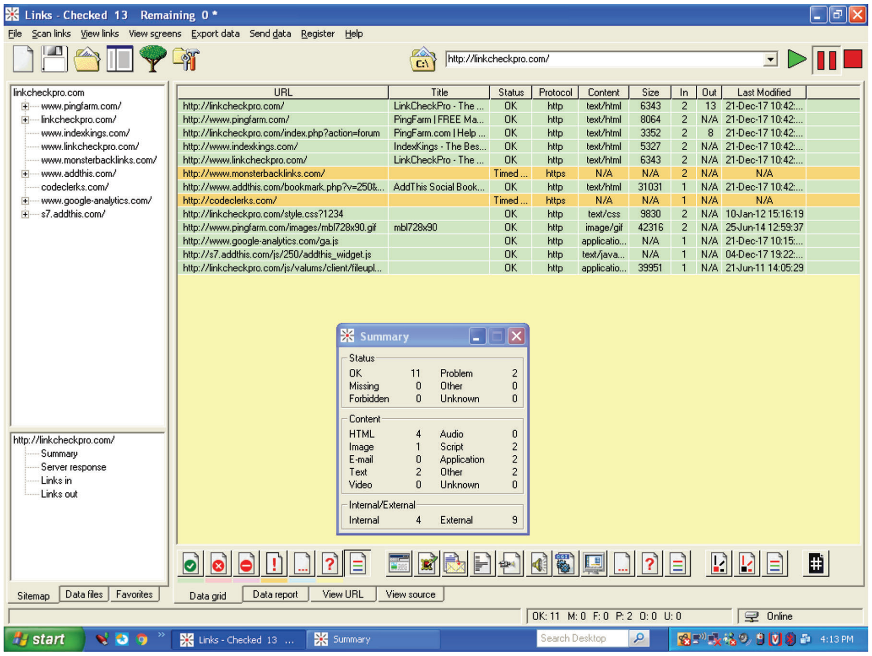
URL	Support Count	Confidence	Relative Confidence
$\{a, b\} \Rightarrow d$	3	0.75	0.25
$\{a, b\} \Rightarrow e$	4	1	0.25
$\{a, c\} \Rightarrow e$	2	1	0.25
$\{a, d\} \Rightarrow e$	3	1	0.25

$\{a, c\} \Rightarrow e$, $\{a, d\} \Rightarrow e$ URLs has more chance to get affected by watering hole attack. Finally proposed method identifies the www.incods2017.in as most frequently used web site by victims. In order to check the availability of watering hole attack on that particular website, a bit of analysis is required as follows in Section 5.

5 Link Analysis

After identifying the frequently used website of target group, a formal verification is done on all internal link and external links. The URL of all internal and external link of particular website is computed using any link analysis tool. In this work, we have used Link analysis pro tool 3.3.37 to analysis the web site [14].

Figure 2 shows the link analysis of a particular website. Totally 13 sub links are available. While observing the URL name of each link it is easy to find out the phishing URL. The phishing URL always differ from the other sub URL which is created by website administrator. For example phishing URL starts with http://yyy.com where as all other starts with https://yyy.com. These kinds of changes are only possible when someone inserts unwanted code or HTML page through which the targeted group secret information is stolen.



URL	Title	Status	Protocol	Content	Size	In	Out	Last Modified
http://linkcheckpro.com/	LinkCheckPro - The ...	OK	http	text/html	6343	2	13	21-Dec-17 10:42:...
http://www.pingfarm.com/	PingFarm | FREE Ma...	OK	http	text/html	8064	2	N/A	21-Dec-17 10:42:...
http://linkcheckpro.com/index.php?action=forum	PingFarm.com | Help ...	OK	http	text/html	3352	2	8	21-Dec-17 10:42:...
http://www.indexkings.com/	IndexKings - The Bes...	OK	http	text/html	5327	2	N/A	21-Dec-17 10:42:...
http://www.linkcheckpro.com/	LinkCheckPro - The ...	OK	http	text/html	6343	2	N/A	21-Dec-17 10:42:...
http://www.monsterbacklinks.com/		Timed ...	https	N/A	N/A	2	N/A	N/A
http://www.addthis.com/bookmark.php?v=250&...	AddThis Social Book...	OK	http	text/html	31031	1	N/A	21-Dec-17 10:42:...
http://codeclerks.com/		Timed ...	https	N/A	N/A	1	N/A	N/A
http://linkcheckpro.com/style.css?1234		OK	http	text/css	9830	2	N/A	10-Jan-12 15:16:19
http://www.pingfarm.com/images/mbl728x90.gif	mbl728x90	OK	http	image/gif	42316	2	N/A	25-Jun-14 12:59:37
http://www.google-analytics.com/ga.js		OK	http	applicatio...	N/A	1	N/A	21-Dec-17 10:15:...
http://s7.addthis.com/js/250/addthis_widget.js		OK	http	text/java...	N/A	1	N/A	04-Dec-17 19:22:...
http://linkcheckpro.com/js/valums/client/fileupl...		OK	http	applicatio...	39951	1	N/A	21-Jun-11 14:05:29

Summary

Status

OK	11	Problem	2
Missing	0	Other	0
Forbidden	0	Unknown	0

Content

HTML	4	Audio	0
Image	1	Script	2
E-mail	0	Application	2
Text	2	Other	2
Video	0	Unknown	0

Internal/External

Internal	4	External	9

Figure 2 Link analysis of a web site.

6 Conclusion

The attacks are the major threat to internet. One among such attacks is called digital waterhole attack. The secret information of an individual/group is easily accessed by the culprits using watering hole attack. To prevent this type of attack, as novelty, we have applied sequential pattern on a browsing history of an individual/group to identify most frequently used website of that target. In addition, link analysis is performed on that identified website to know the URL of each and every link. When any URL of a website differs from other URL's means then alarming the user about watering hole attack.

Acknowledgement

The author would like to thank the management of Kalasalingam Academy of Research and Education for supporting the research through University Research Fellowship and also thank to National Cyber Defense Research Centre for supporting laboratory facility for research.

References

[1] Kirda, E., and Kruegel, C. (2005). Protecting users against phishing attacks with antiphish. In *Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International*, 1, 517–524. IEEE.

[2] Fette, I., Sadeh, N., and Tomasic, A. (2007). Learning to detect phishing emails. In *Proceedings of the 16th international conference on World Wide Web* 649–656. ACM.

[3] Barraclough, P., and Sexton, G. (2015). Phishing website detection fuzzy system modelling. In *Science and Information Conference (SAI), 2015* 1384–1386. IEEE. London, UK, DOI:10.1109/SAI.2015.7237323

[4] Zhang, J., Ou, Y., Li, D., and Xin, Y. (2012). A Prior-based Transfer Learning Method for the Phishing Detection. *JNW*, *7*(8), 1201–1207.

[5] Afroz, S., and Greenstadt, R. (2011, September). Phishzoo: Detecting phishing websites by looking at them. In *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference on* 368–375. IEEE. DOI:10.1109/ICSC.2011.27

[6] Mahmood Ali, M., and Rajamani, L. (2012). "APD: ARM Deceptive Phishing Detector System Phishing Detection in Instant Messengers Using Data Mining Approach", In *Global Trends in Computing and Communication Systems* CCIS 269, 490–502.

[7] Hamid, I. R. A., Abawajy, J., and Kim, T. H. (2013). Using feature selection and classification scheme for automating phishing email detection. *Studies in Informatics and Control*, *22*(1), 61–70.

[8] Na, S. Y., Kim, H., and Lee, D. H. (2014). Prevention schemes against phishing attacks on internet banking systems. *International Journal of Advances in Soft Computing & Its Applications*, *6*(1), 1–5.

[9] Subburaj, T., Suthendran, K., and Arumugam, S. (2017). "Statistical Approach to Trace the Source of Attack Based on the Variability in Data Flows", In *International Conference on Theoretical Computer Science and Discrete Mathematics*, 392–400.

[10] Subburaj T., and Suthendran, K. (2017). "Detection and Trace Back of DDoS Attack Based on Statistical Approach", *Journal of Advanced Research in Dynamical & Control Systems*, 66–74.

[11] Available at: https://krebsonsecurity.com/tag/watering-hole-attack/

[12] Available at: http://www.symantec.com/connect/blogs/internet-explorer-zero-day-used-watering-hole-attack-qa

[13] Available at: https://blog.pivotal.io/data-science-pivotal/case-studies/sequential-pattern-mining-approach-for-watering-hole-attack-detection-2
[14] Available at: http://www.link-checker-pro.com/
[15] Council on Foreign Relations Website Hit by Watering Hole Attack, "IE Zero-Day Exploit". Threat posts the first stop for security news. 2012-12-29. Retrieved 2017-04-02.
[16] Attackers target dozens of global banks with new malware. Symantec Security Response. Retrieved 2017-04-02.l
[17] Available at: https://threatpost.com/researchers-find-blackenergy-apt-links-in-expetr-code/126662

Biographies

T. Subburaj is a Research Scholar in the Department of Computer Applications, Kalasalingam Academy of Research and Education, Krishnankoil, Tamilnadu, India, from 2016. He received his B. Sc in Computer science from Madurai Kamaraj University in 2003; his degree of Master of Computer Applications from Anna University in 2006 and his M.E. Computer Science and Engineering from Anna University in 2012. His current research areas include Distributed system and network security.

Suthendran Kannan received his B.E. Electronics and Communication Engineering from Madurai Kamaraj University in 2002; his M.E. Communication Systems from Anna University in 2006 and his Ph.D Electronics and Communication Engineering from Kalasalingam University in 2015. He was a Research and Development Engineer at Matrixview Technologies Private Limited, Chennai for a couple of years. He is now the Head, Cyber Forensics Research Laboratory and Associate Professor in Information Technology, Kalasalingam Academy of Research and Education. His current research interests include Cyber Security, Communication System, Signal Processing, Image Processing, etc.