Comparison of Phishing Detection Techniques

Parth Parmar
Computer Science & Engineering Department
L.D. Institute of Technology
Ahmedabad, India

Kalpesh Patel
Computer Science & Engineering Department
L.D. Institute of Technology
Ahmedabad, India

***Abstract*—Email has become a popular topic of discussion in today's world. Each month, more & more attacks are launched at the purpose of making web-users believe that they are dealing with a trusted & reliable entity for the purpose of stealing logon credentials, account information and identity information.This type of attack is known as 'phishing'. In this paper we will study about overview of phishing detection and approaches used in respective techniques. This study will help us to build much more strong and robust technique for detection of phished emails by combining multiple techniques and getting a better result.**

Keywords*— *Phishing, email, phishing detection, comparison.

I. INTRODUCTION

According to the survey of Radicati group from April 2010,there are about 1.9 billion users of email worldwide [1]. A 2012 global study reports that 556 million victims per year due to cyber crimes and one of the reasons could be 44% of adult access e-mails via free or unsecured Wi-Fi connections[2].Phishing is a worldwide problem which creates a great effect on both buisness and consumers.The number of worldwide email accounts is expected to increase from an installed base of 3.1 billion in 2011 to nearly 4.1 billion by year-end 2015. This represents an average annual growth rate of 7% over the next four years [3]. It aims at exploiting the weakness in the users. For example, as evaluated in [4], end-users failed to detect 29% of phishing attacks even when trained with the best performing user awareness program.

Due to vast and broad nature of phishing problem,this detection of phished emails study begins by :

- Defining a phishing problem.
- Life Cycle of phishing campaign.
- Detection Approaches.
- Learned lessons from above approaches.

Payment Services continued to be the most-targeted industry sector. Most sectors remained consistent with the first quarter of 2013, except for computer and online gaming, which experienced a notable drop from 5.66 percent in Q1 2013 to 2.03 percent in Q2 2013.

Fig.1 Statistical Highlights for 2nd Quarter2013 to APWG. Source: [5]

II. DEFINING A PHISHING PROBLEM

The Definition of phishing is not consistent as phishing problem is broad and discusses about various scenarios .For example, according to PhishTank:

"Phishing is a fraudulent attempt, normally made through email, to steal your personal information"

But this definition limits the phishing attacks as it is not only concerned with stealing the personal information. For example, a message can tempt the victim to install a script which would in turn transfer the money to the attacker's account, without the need to steal the personal information.

Another definition is provided by Colin Whittaker e. al.[6]:

"We define a phishing page as any web page that, without permission, alleges to act on behalf of a third party with the intention of confusing viewers into performing an action with which the viewer would only trust a true agent of the third party"

III. LIFE CYCLE OF PHISHING CAMPAIGN

An e-mail campaign is a unique e-mail sent out to multiple number of users, redirecting them to a specific phishing web site .

	April	May	June
Number of unique phishing websites detected	36,480	44,511	38,110
Number of unique phishing e-mail reports (campaigns) received by APWG from consumers	20,086	18,297	14,698
Number of brands targeted by phishing campaigns	441	431	425
Country hosting the most phishing websites	USA	USA	USA
Contain some form of target name in URL	50.92%	57.45%	51.52%
No hostname; just IP address	4.57%	5.23%	5.26%
Percentage of sites not using port 80	0.38%	0.45%	0.80%

Fig.2 Statistical Highlights for 2nd Quarter2013 to APWG. Source: [5]

Whenever a phishing campaign is started (e.g. by sending phishing emails to users), the first security measure is to detect the campaign. The detection techniques are large enough and could involve techniques used by service providers to detect the attacks, user awareness programs, and end-user client software classification.

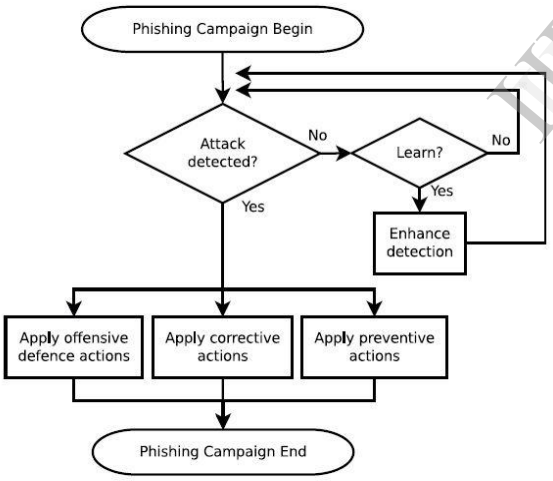

Fig.3 The life-cycle of phishing campaigns from the perspective of anti-phishing techniques. Source :[7]

The ability of detecting phishing campaigns can be enhanced more whenever a phishing campaign is detected through learning from such experience. For example, by learning from previous phishing campaigns, it is possible to improve the detection of future phishing campaigns. Such learning can be performed by either a software, or human observer(who is well-known about identifying phishing campaign).

Once the phishing campaign is detected actions like offensive defense, correction and prevention can be applied. However, if the phishing campaign is not detected , then none of the above actions can be applied. This focuses on the importance of the detection phase.

IV. DETECTION APPROACHES

The detection solutions of phished emails can be basically classified into two types:

1. User training approaches : End-users can be educated to understand the nature of the phishing attack. This is contrary to the categorization in [8] where user training was considered a preventative approach.
2. Software classification approaches : Build automated software classifiers.

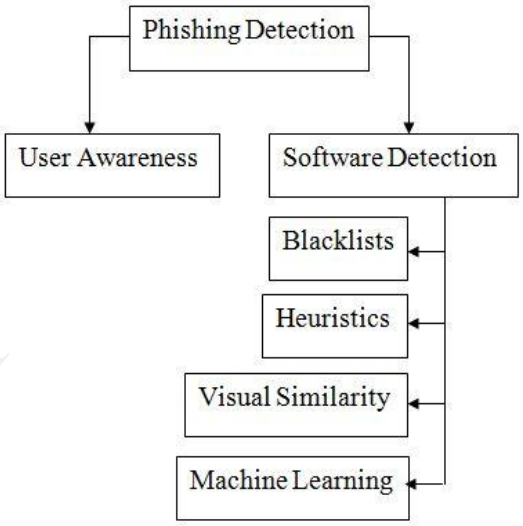

Fig.4 Overview of phishing detection approaches.

TABLE I
COMPARISON THE ADVANTAGES AND DISADVANTAGES BETWEEN PHISHING DETECTION TECHNIQUES

Detection Technique	**Advantages**	**Disadvantages**
Blacklists	-Requiring low resources on host machine -Effective when minimal FP rates are required.	-Mitigation of zero-hour phishing attacks. -Can result in excessive queries with heavily loaded servers.
Heuristics and visual similarity	-Mitigate zero-hour attacks.	-Higher FP rate than blacklists. -High computatio-nal cost.
Machine Learning	-Mitigate zero-hour attacks. -Constuct own classification models.	-Time consuming. -Costly. -Huge number of rules.

V. SUMMARY

The Comparison of techniques has shown that Machine Learning techniques are most promising, but it also has the disadvantages of greater computational cost. New techniques can be developed for having low false positives by combining blacklists and heuristics approaches. As a future work, phishing detection techniques from the perspective of thier computational cost and energy consumption can be thought of.

REFERENCES

1. http://www.radicati.com
2. N Sridhar,D.L. Bhaskari and P.S. Avadhani; Inverted Pyramid Approach for E-Mail Forensics Using Heterogeneous Forensics Tools. In CSI Communications(July 2013) 21-23.
3. Sara Radicati Email Statistics Report,2011-2015
4. S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs,"Who falls for phish?: a demographic analysis of phishing susceptibility and effectiveness of interventions," in *Proceedings of the 28th international conference on Human factors in computing systems*, ser. CHI '10.New York, NY, USA: ACM, 2010, pp. 373–382.
5. APWG ,Phishing Activity trends reports, April-June 2013.
6. C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," in *NDSS '10*, 2010.
7. Khonji, Mahmoud; Iraqi, Youssef; Jones, Andrew, "Phishing Detection: A Literature Survey," *Communications Surveys & Tutorials, IEEE* , vol.15, no.4, pp.2091,2121, Fourth Quarter 2013.
8. S. Abu-Nimeh, D. Nappa, X. Wang, and S. Nair, "A comparison of machine learning techniques for phishing detection," in *Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit*, ser. eCrime '07. New York, NY, USA: ACM, 2007, pp. 60– 69.