Computer Science Review 29 (2018) 44–55

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

A recent review of conventional vs. automated cybersecurity anti-phishing techniques

Issa Qabajeh [b], Fadi Thabtah [a,*], Francisco Chiclana [b]

[a] Digital Technology Department, Manukau Institute of Technology, Auckland, New Zealand
[b] Centre for Computational Intelligence, De Montfort University, Leicester, UK

ARTICLE INFO

Article history:
Received 12 September 2017
Received in revised form 23 May 2018
Accepted 28 May 2018

Keywords:
Classification
Computer security
Phishing
Machine learning
Web security
Security awareness

ABSTRACT

In the era of electronic and mobile commerce, massive numbers of financial transactions are conducted online on daily basis, which created potential fraudulent opportunities. A common fraudulent activity that involves creating a replica of a trustful website to deceive users and illegally obtain their credentials is website phishing. Website phishing is a serious online fraud, costing banks, online users, governments, and other organisations severe financial damages. One conventional approach to combat phishing is to raise awareness and educate novice users on the different tactics utilised by phishers by conducting periodic training or workshops. However, this approach has been criticised of being not cost effective as phishing tactics are constantly changing besides it may require high operational cost. Another anti-phishing approach is to legislate or amend existing cyber security laws that persecute online fraudsters without minimising its severity. A more promising anti-phishing approach is to prevent phishing attacks using intelligent machine learning (ML) technology. Using this technology, a classification system is integrated in the browser in which it will detect phishing activities and communicate these with the end user. This paper reviews and critically analyses legal, training, educational and intelligent anti-phishing approaches. More importantly, ways to combat phishing by intelligent and conventional are highlighted, besides revealing these approaches differences, similarities and positive and negative aspects from the user and performance prospective. Different stakeholders such as computer security experts, researchers in web security as well as business owners may likely benefit from this review on website phishing.

Contents

* Corresponding author.
E-mail addresses: P12047781@myemail.dmu.ac.uk (I. Qabajeh), fadi.fayez@manukau.ac.nz (F. Thabtah), chiclana@dmu.ac.uk (F. Chiclana).

https://doi.org/10.1016/j.cosrev.2018.05.003
1574-0137/

1. Introduction

With the advanced development of computer hardware, especially computer networks and cloud technology services, online and mobile commerce have significantly increased in the last few years [1]. Indeed, the number of customers who perform online purchase transactions has dramatically increased and large monetary values are daily exchanged through electronic means, such as private payment gateways, that are usually verified by secure socket layer (SSL) [2]. Despite the convenience associated with online transactions from both user and business prospectives, an online threat has emerged: phishing.

Phishing attacks are attempts to access online users' sensitive financial information using fake websites that are visually similar to authentic websites [3]. In phishing attacks, social engineering techniques are normally utilised to redirected users to the malicious website. Specifically, an email is sent to users from apparent trustworthy sources, urging them to adjust their login information by clicking/following a hyper link [4]. Phishing techniques include spear phishing, which is a focused attack in which emails are sent to employees of a business in an attempt to access a company's computer system, or whaling, that targets senior corporate executives [5]. Unfortunately, the consequences of phishing are fatal because affected legitimate users become vulnerable to identity theft and information breach and no longer trust online commerce and electronic banking [6]. For instance, Gartner Group [7] publishes periodic reports that revealed financial damages caused by phishing attacks. In addition, to raise awareness about phishing an international body that aims to minimise online threats including pharming, spoofing, phishing and malware, the Anti-Phishing Work Group (APWG), was created [8]. APWG periodically disseminates reports for the online community on recent cyber-attacks, with a recent report stating the rapid increase of phishing websites to 17,000 in the month of December 2014 alone [9]. A recent report published by APWG revealed that there were approximately 1,220,523 phishing attacks in 2016.

It seems imperative that users, as well as businesses, adopt renewable anti-phishing tools or strategies to reduce phishing activities and protect themselves from their potential negative impacts. This is important because phishing attacks are constantly changing and new deceptions are emerging all the time. Anti-phishing solutions adopting DM (ML) are shown to be more practical and effective in combating phishing because they work automatically and are capable of revealing concealed knowledge that online users are not aware of, especially with respect to the relationship among website features and phishing activities. This hidden knowledge, when combined with human experience, can result in an effective shield for protecting users from phishing (add a reference).

In this paper, we investigate the phishing problem and define it in a classification ML context. We then discuss common, traditional, strategies in addition to computerised techniques developed to combat phishing. More importantly, the paper thoroughly investigates traditional and ML anti-phishing classification techniques and critically analyses their benefits and disadvantages theoretically. There have been few former reviews on phishing such as Suganya [10], Mohammad et al. [11,12], Sahu and Dubey [13], Almomani et al. [14] and Basnet et al. [15] among others. For instance, Almomani et al. [14] reviewed a number of filtering techniques to combat phishing. The authors have focused only on technical solutions of detecting phishing emails by reviewing techniques related to Bag of Words, frequency analysis, blacklists, support vector machines and other artificial intelligence search methods. Little information concerning non-technical solution were provided. Instead, the authors paid full attention to review automated solutions that can be integrated within email systems to detect phishing attacks. Lastly, the authors reported different disseminated research results in a table format to show the performance of various different machine learning techniques against email phishing data. However, it will be hard to generalise such performance due to the fact that these results have been derived from datasets with different characteristics. Overall, the survey provided was insightful and it provide rich information to users in order to reduce the chance of falling into email phishing attacks.

Gupta et al. [16] reviewed different types of phishing attacks and then discussed a number of anti-phishing approaches including social engineering ones. More importantly, the authors showed features related to phishing attacks that have been collected from previous research works including [14,17,18] and [32] among others. Lastly, the authors highlighted emergent trends in phishing and modern technologies such as the Internet of Things. Gupta et al. [19] highlighted recent challenges and new emergent trends in phishing attacks. The focus of the researcher was on the new technology of the Intent of Things and spear phishing. The authors also discussed recent phishing datasets and their features.

Most of phishing reviews have covered partly one or more of phishing aspects. For instance, Suganya [10] and Sahu and Dubey [13] briefly reviewed phishing attacks without showing the ways to combat them or their pros and cons. Mohammad et al. [11,12] discussed in general common solutions of website phishing without providing grounds for recommendations besides not covering specific intelligent approaches. Almomani et al. [14] reviewed intelligent solutions to detect phishing emails. Lastly, Basnet et al. [15] compared only few intelligent anti-phishing solutions without on elaborating the other computerised and classic approaches of anti-phishing. Therefore, this article not only comprehensively reviews phishing from wider prospective but also it critically analyses traditional and automated anti-phishing solutions.

This paper serves researchers, organisations' managers, computer security experts, lecturers, and students who are interested in understanding phishing and its corresponding intelligent solutions. This is since wider potential solutions have been critically analysed and experimentally compared besides presenting classic solutions including educational, legal, and software based. This paper is structured as follows: Section 2 presents the phishing problem, its history, and its lifecycle. Section 3 critically analyses common classic methods of combating phishing besides critically analysing them. Section 4 is devoted to intelligent anti-phishing solutions that employ different strategies in deriving the anti-phishing models. Section 5 provides the conclusions.

2. Phishing background

2.1. Phishing history

Phishing comes from the word "fishing", in which the phisher throws a bait and awaits for potential users to take a bite. Phishing is not recent as an online risk, with its origin rooted in a social engineering method using telephones known as "phone phreaking" [20]. It was during the 1990s period when the internet community started to grow that phishing was originally observed as an online threat, especially in the United States [15].