Information & Management 51 (2014) 845–853

Contents lists available at ScienceDirect

Information & Management

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

A domain-feature enhanced classification model for the detection of Chinese phishing e-Business websites

Dongsong Zhang [a,b], Zhijun Yan [b,*], Hansi Jiang [b], Taeha Kim [c,**]

[a] Department of Information Systems, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA
[b] School of Management & Economics, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China
[c] Department of Management Information Systems, College of Business & Economics, Chung-Ang University, 84 HeukSeok-Ro, Dongjak-Gu, Seoul, Korea 156-756

ARTICLE INFO

Article history:
Received 1 November 2013
Received in revised form 24 May 2014
Accepted 2 August 2014
Available online 12 August 2014

Keywords:
Phishing websites
E-business
Classification
Detection
Feature vectors

ABSTRACT

We propose a novel classification model that consists of features of website URLs and content for automatically detecting Chinese phishing e-Business websites. The model incorporates several unique domain-specific features of Chinese e-Business websites. We evaluated the proposed model using four different classification algorithms and approximately 3,000 Chinese e-Business websites. The results show that the Sequential Minimal Optimization (SMO) algorithm performs the best. The proposed model outperforms two baseline models in detection precision, recall, and F-measure. The results of a sensitivity analysis demonstrate that domain-specific features have the most significant impact on the detection of Chinese phishing e-Business websites.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

The past decade has seen a remarkable growth of e-Business worldwide. The convenience of e-Business, however, exposes consumers to a variety of security and privacy concerns. Among them, phishing, a form of online identity theft associated with both social engineering and technical subterfuge, is a major threat. Phishing is best understood as distinct methods identity thieves use to illegally obtain online users' personal information by enticing unwitting users to give out their identity or financial information either unknowingly or under false impressions or by deceiving users to allow unauthorized access to their computers and personal data.

There has been an increasing number of phishing e-Business websites that aim to acquire consumers' personal and sensitive information illegally for financial gains or to mislead consumers into conducting business transactions that will never be fulfilled by masquerading a phishing website as a trustworthy e-Business website [19]. Phishing e-Business websites pose a considerable threat not only to e-Business but also to Internet security and consumer privacy. According to APWG's (Anti-Phishing Working Group) Phishing Activity Trends Report for the first quarter of 2012 [6], there were 56,859 unique phishing websites detected in February alone. Phishing e-Business websites seriously affect the development of online financial services and e-Commerce, endanger public interests, and negatively affect public interests in e-Commerce. Therefore, developing effective approaches to detecting phishing e-Business websites is critical to mitigating this threat and its associated financial losses [12].

There are two common types of phishing attacks in e-Business. One is to create fake e-Business websites that look very similar to real, authentic e-Business websites in terms of domain names and Web content. Once the deceived consumers mistakenly log into a phishing website, their user name and password can be stolen and used by criminals to log into the authentic website for illegal financial gain. These phishing websites (e.g., the left website in Fig. 1) are called spoof sites.

Another type of phishing website is referred to as a concocted site, which is a phishing website without an authentic counterpart. For example, criminals may post fake product sales information on fake websites, then disappear after receiving the consumer's payment. Although public awareness of phishing websites has been steadily increasing over the years, the number of phishing websites and the resultant damage have grown at an even faster

* Corresponding author at: School of Management & Economics, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China. Tel.: 86 10 68912845.
** Corresponding author at: College of Business and Economics, Chung-Ang University, 84 HeukSeok-Ro, Dongjak-Gu, Seoul, Korea 156-756. Tel.: 82 2 820 5543.
E-mail addresses: zhangd@umbc.edu (D. Zhang), yanzhijun@bit.edu.cn (Z. Yan), jianghansi@qq.com (H. Jiang), tkim@cau.ac.kr (T. Kim).

http://dx.doi.org/10.1016/j.im.2014.08.003
0378-7206/© 2014 Elsevier B.V. All rights reserved.

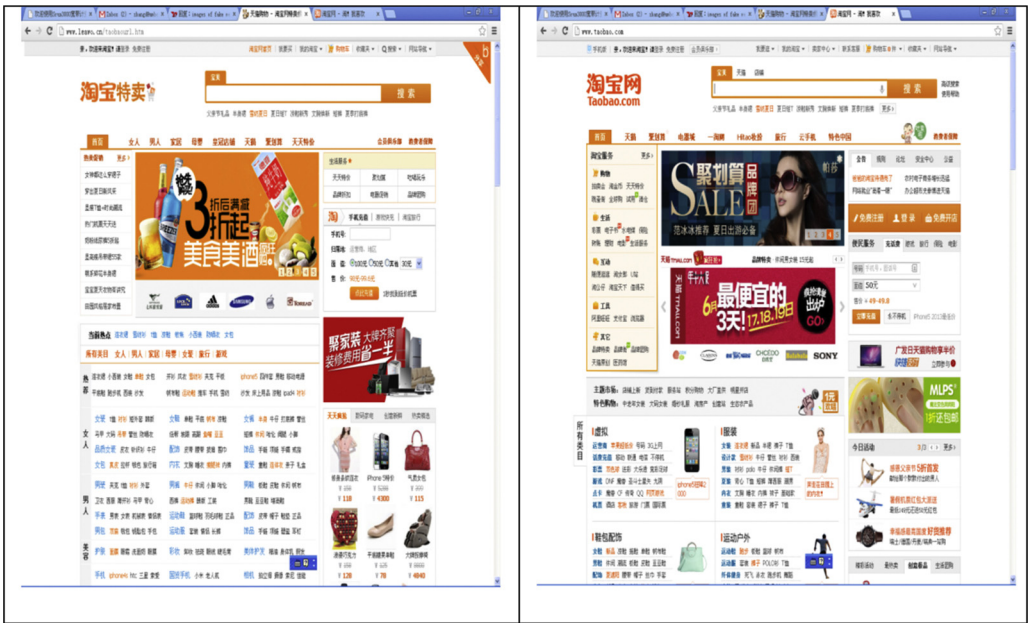

Fig. 1. Fake (left) versus authentic (right) Chinese e-Business websites.

pace. According to the Anti-Phishing Alliance of China, 24,535 new Chinese phishing websites emerged in 2012, and 60 million Chinese online users became victims of phishing websites and suffered a loss of US$4.64 billion between July 2011 and June 2012. More than 60% of Chinese phishing websites have been e-Business websites.

Detecting phishing e-Business websites successfully is a major goal of anti-phishing, but it is a challenging task due to the authentic appearance of phishing websites. Phishing websites are often professional-looking and sophisticated in terms of design and appearance, making their detection difficult [2,25,37]. In Jagatic et al.'s [23] and Wu et al.'s [39] studies, 60% and 72% of participants provided personal information to fake websites, respectively, due to their inability to identify them as phishing websites.

There have been a variety of approaches to phishing website detection. Some approaches are based on the recognition of the content and URL of a website (e.g., [18,21,34]); some segment a website into images and then analyze those images (e.g., [10,14,42]); and others use third-party search engines (e.g., [22]). However, these existing approaches have various limitations, such as reliance on prior knowledge about authentic websites and the lack of taking unique features of a particular domain into consideration. The vast majority of current approaches are generic approaches to phishing website detection instead of domain-specific approaches.

Compared with general phishing websites, there are unique features of Chinese e-Business websites that make existing approaches to general phishing website detection either not applicable or ineffective. For example, some current detection methods examine certain features of a website, such as whether its URL contains keywords such as 'eBay' and 'PayPal', which are rarely included in the URL of Chinese e-Business websites. Second, China's Ministry of Information Industry requires every legitimate e-Business to register the domain name of its website, and the registry of those domain names is accessible by the public. Third, there is special content on Chinese e-Business websites that may not exist in other e-Business websites, let alone general phishing websites. For example, there are several e-Business certificates available in China. Although it is not required, many Chinese e-Businesses post their certificate information on their websites, aiming to gain more trust from online consumers. The examples of domain-specific features described above could be potentially helpful in the detection of phishing Chinese e-Business websites, but they have never been examined. Therefore, existing approaches to generic phishing website detection may not be effective for the detection of Chinese phishing e-Business websites.

To cope with the increasing problem of Chinese phishing e-Business websites and address the limitations of existing detection approaches, we propose a new classification model for detecting those websites. The overarching research question of this study is: *Can the incorporation of domain-specific features into a phishing e-Business website detection model improve the detection performance in comparison with a generic detection model?* There are three contributions of this study. First, previous studies on phishing website detection mainly focused on building generic models. As a result, the models developed only included general features of websites. In reality, however, phishing websites in different domains may exhibit different unique characteristics, making generic detection models potentially less effective. There has been little research on phishing website detection for a specific domain. In the proposed model, we incorporate domain-specific features that reflect unique characteristics of Chinese e-Business websites in addition to some generic website features adopted from previous research. Our model neither requires user expertise and prior knowledge about authentic websites nor consults centralized whitelists or blacklists to determine whether a target website is a phishing website. It can be used to detect both types of phishing attacks introduced earlier. Second, although China has witnessed the fastest growth of phishing e-Business websites and has suffered formidable financial losses attributable to those websites, there has been little research on how to build effective models for detecting those websites. In this study, we have built and empirically evaluated the proposed model with approximately three thousand authentic and phishing Chinese e-Business websites using four different machine learning algorithms. Third, we conducted a sensitivity analysis on the influence of individual predictive features in the model on detection performance to identify the most influential features in order to make the model more parsimonious, which has rarely been done in previous studies. Such an evaluation approach offers new insights for further improving a detection model.

The rest of the paper will be organized as follows. Section 2 introduces related work on the automated detection of phishing

Download English Version:

https://daneshyari.com/en/article/555539

Download Persian Version:

https://daneshyari.com/article/555539

Daneshyari.com