Computers & Security 95 (2020) 101855

Contents lists available at ScienceDirect

Computers & Security

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

LogoSENSE: A companion HOG based logo detection scheme for phishing web page and E-mail brand recognition

Ahmet Selman Bozkir, Murat Aydos

Department of Computer Engineering, Hacettepe University, Ankara, TURKEY

ARTICLE INFO

Article history:
Received 25 December 2019
Revised 5 April 2020
Accepted 21 April 2020
Available online 8 May 2020

Keywords:
Phishing
Logo Detection
Computer Vision
HOG
Machine Learning

ABSTRACT

With the advent of Internet and opportunities in e-commerce, a visual perception oriented cyber-attack so-called phishing has become one of the tremendous problems of the cyber world since it aims to access user credentials in order to gain illegal financial profit and steal sensitive personal data. In order to fight with this security threat, various studies using a different source of information such as URL, text content, DOM trees or visual features belonging to web pages have been utilized. Apart from other works, we propose a companion scheme to recognize brands of “zero hour” phishing web pages by localizing and classifying the target brand logos involved in page screenshots by solely use of computer vision methods in object detection manner. For this purpose, the features of Histogram of Oriented Gradients (HOG) have been employed to obtain visual representations of target brand logos in scale invariant fashion. In addition, throughout the classification, a max-margin loss equipped SVM classifier has been used in order to work with a low number of training images and to decrease the number of false positives. Moreover, we prepared a publicly available dataset having a total of 3060 training and 1979 unique phishing and legitimate web page/e-mail snapshots along with their bounding box annotations for evaluation and further academic usage. Detailed experiments show that, at the best configuration, our schema named “LogoSENSE” is able to achieve 93.50% precision and of 77.94% recall score along with obtaining F1 score of 85.02%. The experiments show that the proposed approach outperforms SIFT based detection and presents comparative results against a state-of-art deep learning based object detection method. As a result, LogoSENSE serves promising results in terms of detection accuracy and run-time efficiency, yielding a companion tool that can be used as a brand recognition mechanism for phishing web pages and emails.

1. Introduction

Significant progress in e-commerce and the growing number of online services (e.g., e-mail, cloud data, online payment systems) have eased not only life but also led an increase in a special kind of cybercrime named phishing. By definition, phishing is a security threat that attempts to deceive innocent users into capturing their confidential information (e.g., username, password, credit card details and social security number) by combining social engineering and web site spoofing techniques (Chiew et al., 2015; Rao and Ali, 2015). The lifecycle of a typical phishing attack starts with receiving a fake e-mail, SMS or instant message from scammers which attempts to make users think and believe that it is coming from a legitimate source. It should be noted that these messages usually involve compelling statements and a link pointing to scammer’s fake web page that mimics the target brand’s legitimate web page. Once the user inputs his/her credential, the life cycle of the attack eventually ends by sending the sensitive information to phishers in order to be used for various purposes such as online fraud or exploit of private data.

According to the phishing trend reports of Anti-Phishing Working Group (APWG, 2018), the total financial losses due to the attacks in 2014 and 2015 were 4.5 and 4.6 billion dollars respectively (Jain and Gupta, 2017). Moreover, as of 2013, approximately 450.000 phishing attacks were recorded. Apart from financial losses, in recent years, APWG has also pointed out a new emerging threat, so-called spear-phishing – a variant of phishing – which targets specific users or companies to obtain their private contents (e.g. business secrets, personal repositories) instead of financial profit. Further, APWG’s Q3 (third quarter) report of 2017 highlights the increasing number of phishing attacks targeting the cryptocurrency sector.

Although it has been studied for almost two decades, phishing is still not a fully solved problem since there exists an arms race

E-mail address: selman@cs.hacettepe.edu.tr (A.S. Bozkir).

https://doi.org/10.1016/j.cose.2020.101855
0167-4048/© 2020 Elsevier Ltd. All rights reserved.