Web Phishing Detection
Literature Survey

Submitted By,

Gowri Shankar.m

Gokul .R

Karthick .k

Jaffarson .JS

Janagan

Paper 1 Web phishing detection techniques: a survey on the state-of-the-art, taxonomy and future directions

Author:

M. Vijayalakshmi,S. Mercy Shalinie,Ming Hour Yang,Raja Meenakshi U.

Published:

23 Septembar 2020,

Internet dragged more than half of the world's population into the cyber world. Unfortunately, with the increase in internet transactions, cybercrimes also increase rapidly. With the anonymous structure of the internet, attackers attempt to deceive the end-users through different forms namely phishing, malware, SQL injection, man-in-the-middle, domain name system tunnelling, ransomware, web trojan, and so on. Amongst them, phishing is the most deceiving attack, which exploits the vulnerabilities in the end-users. Phishing is often done through

emails and malicious websites to lure the user by posing themselves as a trusted entity. Security experts have been proposing many anti-phishing techniques. Till today there is no single solution that is capable of mitigating all the vulnerabilities. A systematic review of current trends in web phishing detection techniques is carried out and a taxonomy of automated web phishing detection is presented. The objective of this study is to acknowledge the status of current research in automated web phishing detection and evaluate their performance. This study also discusses the research avenues for future investigation.

Paper 2: Web Phishing Detection Using a Deep Learning Framework

Author:

Yuxiang Guan, Futai Zou, Wei Wang,

Published:

26 September 2016,

Web service is one of the key communications software services for the Internet. Web phishing is one of many security threats to web services on the Internet. Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity. It will lead to

information disclosure and property damage. This paper mainly focuses on applying a deep learning framework to detect phishing websites. This paper first designs two types of features for web phishing: original features and interaction features. A detection model based on Deep Belief Networks (DBN) is then presented. The test using real IP flows from ISP (Internet Service Provider) shows that the detecting model based on DBN can achieve an approximately 90% true positive rate and 0.6% false positive rate.

Web service is a communication protocol and software between two electronic devices over the Internet [1]. Web services extends the World Wide web infrastructure to provide the methods for an electronic device to connect to other electronic devices [2]. Web services are built on top of open communication protocols such as TCP/IP, HTTP, Java, HTML, and XML. Web service is one of the greatest inventions of mankind so far, and it is also the most profound manifestation of computer influence on human beings [3].

Paper 3:Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text

Author:

M.A.Adebowale, K.T.Lwin, E.Sánchez,

Published:

25 April 2018,

A phishing attack is one of the most significant problems faced by online users because of its enormous effect on the online activities performed. In recent years, phishing attacks continue to escalate in frequency, severity and impact. Several solutions, using various methodologies, have been proposed in the literature to counter the web-phishing threats. Notwithstanding, the existing technology cannot detect the new phishing attacks accurately due to the insufficient integration of features of the text, image and frame in the evaluation process. The use of related features of images, frames and text of legitimate and non-legitimate websites and associated artificial intelligence algorithms to develop an integrated method to address these together. This paper presents an Adaptive Neuro-Fuzzy Inference System (ANFIS) based robust scheme using the integrated features of the text, images and frames for web-phishing detection and protection. The proposed solution achieves 98.3% accuracies. To our best knowledge, this is the first work that considers the best-integrated text, image and frame feature based solution for phishing detection scheme.

Paper 4: Systematization of Knowledge (SoK): A Systematic Review of Software-Based Web Phishing Detection

Author:

Zuochao Dou, Issa Khalil, Ala Al-Fuqaha, Mohsen Guizani

Publisher:

Oct 1 2017

Phishing is a form of cyber attack that leverages social engineering approaches and other sophisticated techniques to harvest personal information from users of websites. The average annual growth rate of the number of unique phishing websites detected by the Anti Phishing Working Group is 36.29% for the past six years and 97.36% for the past two years. In the wake of this rise, alleviating phishing attacks has received a growing interest from the cyber security community. Extensive research and development have been conducted to detect phishing attempts based on their unique content, network, and URL characteristics. Existing approaches differ significantly in terms of intuitions, data analysis methods, as well as evaluation methodologies. This warrants a careful systematization so that the advantages and limitations of each approach, as well as the applicability in different contexts, could be analyzed and contrasted in a rigorous and principled way. This paper presents a systematic study of phishing detection schemes, especially software based ones. Starting from the phishing

detection taxonomy, we study evaluation datasets, detection features, detection techniques, and evaluation metrics. Finally, we provide insights that we believe will help guide the development of more effective and efficient phishing detection schemes.

Paper 5: Web Phishing Detection Based on Page Spatial Layout Similarity Author:

, Hongji Yang

Published:

July 8, 2013,

Web phishing is becoming an increasingly severe security threat in the web domain. Effective and efficient phishing detection is very important for protecting web users from loss of sensitive private information and even personal properties. One of the keys of phishing detection is to efficiently search the legitimate web page library and to find those page that are the most similar to a suspicious phishing page. Most existing phishing detection methods are focused on text and/or image features and have paid very limited attention to spatial layout characteristics of web pages. In this paper, we propose a novel phishing detection method that makes use of the informative spatial layout characteristics of web pages. In particular, we develop two different options to extract the spatial layout features as rectangle blocks from a given web page. Given two web pages, with their respective spatial layout features, we propose a page similarity definition that takes into account their spatial layout characteristics. Furthermore, we build an R-tree to index all the

spatial layout features of a legitimate page library. As a result, phishing detection based on the spatial layout feature similarity is facilitated by relevant spatial queries via the R-tree. A series of simulation experiments are conducted to evaluate our proposals. The results demonstrate that the proposed novel phishing detection method is effective and efficient.