WEB PHISHING DETECTION

Team ID: PNT2022TMID50116

Team Member: J.Abila Jesy, R.Anisha, M.AnuPriya, M.Celsiya

Paper-1: OluwatobiAyodejiAkanbi, IrajSadeghAmiri, ElaheFazeldehkordi, A Machine-Learning Approach to Phishing Detection and Defense, 2015

Phishing is one of the most widely-perpetrated forms of cyber attack, used to gather sensitive information such as credit card numbers, bank account numbers, and user logins and passwords, as well as other information entered via a web site. The authors of A Machine-Learning Approach to Phishing Detection and Defense have conducted research to demonstrate how a machine learning algorithm can be used as an effective and efficient tool in detecting phishing websites and designating them as information security threats. This methodology can prove useful to a wide variety of businesses and organizations who are seeking solutions to this long-standing threat. A Machine-Learning Approach to Phishing Detection and Defense also provides information security researchers with a starting point for leveraging the machine algorithm approach as a solution to other information security threats.

Paper-2: Mahar Aburrous, Alamgir Hossain, KeshavDahal, LAP Lambert Academic Publishing Theodor-Heuss-Ring 26, Phishing Website Detection using Intelligent Data Mining Techniques: Design and Development of an Intelligent Association Classification Fuzzy Based Scheme for Phishing Website Detection, January 2012

Detecting phishing website is a complex task which requires significant expert knowledge and experience. So far, various solutions have been proposed and developed to address these problems. Most of these approaches are not able to make a decision dynamically, giving rise to a large number of false positives. This is mainly due to limitation of the previously proposed approaches. In this book, we investigate and develop the application of an intelligent fuzzybased classification system for phishing website detection. The proposed intelligent phishing detection system employed Fuzzy Logic (FL) model with association classification mining algorithms. Different phishing experiments which cover all phishing attacks, motivations and deception behavior techniques have been conducted to cover all phishing concerns. A comparative study and analysis showed that the proposed learning approach has a higher degree of predictive and detective capability than existing models. The proposed system was developed, tested and validated by incorporating the scheme as a web based plug-ins phishing toolbar to provide an effective help for real-time phishing website detection for all internet users.

Paper-3: Shekhar Khandelwal, Rik Das,Phishing Detection Using Content-Based Image Classification,1st Edition, June 2022,New York

Phishing Detection Using Content-Based Image Classification is an invaluable resource for any deep learning and cyber security professional and scholar trying to solve various cyber security tasks using new age technologies like Deep Learning and Computer Vision. With various rule based phishing detection techniques at play which can be bypassed by phishers, this book provides a step-by-step approach to solve this problem using Computer Vision and Deep Learning techniques with significant accuracy. Phishing Detection Using Content-Based Image Classification is an invaluable resource for any deep learning and cyber security professional and scholar trying to solve various cyber security tasks using new age technologies like Deep Learning and Computer Vision. With various rule-based phishing detection techniques at play which can be bypassed by phishers, this book provides a step-by-step approach to solve this problem using Computer Vision and Deep Learning techniques with significant accuracy.
The book offers comprehensive coverage of the most essential topics, including:
Programmatically reading and manipulating image data
Extracting relevant features from images
Building statistical models using image features
Using state-of-the-art Deep Learning models for feature extraction
Build a robust phishing detection tool even with less data
Dimensionality reduction techniques
Class imbalance treatment
Feature Fusion techniques
Building performance metrics for multi-class classification task
Another unique aspect of this book is it comes with a completely reproducible code base developed by the author and shared via python notebooks for quick launch and running capabilities. They can be leveraged for further enhancing the provided models using new advancement in the field of computer vision and more advanced algorithms.

Paper-4: Samuel Marchal∗, Kalle Saari∗, Nidhi Singh†and N. Asokan∗‡∗Aalto University†Intel Security,Know Your Phish: Novel Techniques for Detecting Phishing Sites and their Targets,25 Apr 2016

data are thus very attractive. In this paper, we introduce new approach -Phishing is a major problem on the Web. Apr 2016the significant attention it has received over the years, there has draw backs. Our goal is to identify whether a given webpage been no definitive solution. While the state-of-the-art solutions is a phish, and, if it is, identify the target it is trying to mimic. have reasonably good performance, they require a large amount Our approach is based on two core conjectures: of training data and are not adept at detecting phishing attacks• Modeling phisher limitations: To increase their chances against new targets. In this paper, we begin with two core observations: (a) although of success, phishers try to make their phish mimic its phishers try to make a phishing webpage look similar to its target, target closely and obscure any signal that might tip off the they do not have unlimited freedom in structuring the phishing victim. However, in crafting the structure of the phishing webpage; and (b) a webpage can be characterized by a small set webpage, phishers are restricted in two significant ways. of key terms; how these key terms are used in different partsFirst, external hyperlinks in the phishing webpage, espe of a webpage is different in the case of legitimate and phishing webpages. Based on these observations, we develop a phishingcially those pointing to the target, are to domains outside detection system with several notable properties: it requires the control of phishers. Second, while phishers can freely very little training data, scales well to much larger test data, change most parts of the phishing page, the latter part is language independent, fast, resilient to adaptive attacks and of its domain name is constrained as they are limited implemented entirely on client-side. In addition, we developed to domains that the phishers control. We conjecture that a target identification component that can identify the target website that a phishing webpage is attempting to mimic. The by modeling these limitations in our phishing detection target detection component is faster than previously reported classifier, we can improve its effectiveness. systems and can help minimize false positives in our phishing• Measuring consistency in term usage: A webpage can detection system