Web Phishing Detection

Abstract

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.

Introduction

Web service is a communication protocol and software between two electronic devices over the Internet. Web services extend the World Wide web infrastructure to provide the methods for an electronic device to connect to other electronic devices. 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.

With the rapid development of the Internet and the increasing popularity of electronic payment in web service, Internet fraud and web security have gradually been the main concern of the public [4]. Web Phishing is a way of such fraud, which uses social engineering techniques through short messages, emails, and WeChat [5] to induce users to visit fake websites to get sensitive information like their private account, token for payment, credit card information, and so on.

The first phishing attack on AOL (America Online) can be traced back to early 1995 [6]. A phisher successfully obtained AOL users personal information. It may lead to not only the abuse of credit card information, but also an attack on the online payment system entirely feasible.

The phishing activity in early 2016 was the highest ever recorded since it began monitoring in 2004. The total number of phishing attacks in 2016 was 1,220,523. This was a 65 percent increase over 2015. In the fourth quarter of 2004, there were 1,609 phishing attacks per month. In the fourth quarter of 2016, there was an average of 92,564 phishing attacks per month, an increase of 5,753% over 12 years.

Implementation :

Phishing is a form of fraudulent attack where the attacker tries to gain sensitive information by posing as a reputable source. In a typical phishing attack, a victim opens a compromised link that poses as a credible website. The victim is then asked to enter their credentials, but since it is a “fake” website, the sensitive information is routed to the hacker and the victim gets “hacked.”

Phishing is popular since it is a low effort, high reward attack. Most modern web browsers, antivirus software and email clients are pretty good at detecting phishing websites at the source, helping to prevent attacks. To understand how they work, this blog post will walk you through a tutorial that shows you how to build your own phishing URL detection using Python and machine learning:

1. **Identify the criteria** that can recognize fake URLs
2. **Build a decision tree** that can iterate through the criteria
3. **Train our model** to recognize fake vs real URLs
4. **Evaluate our model** to see how it performs
5. **Check for false positives/negatives**

We will focus on the detection model using a deep learning framework. The main contributions are as follows:

(i) We present two feature types for web phishing detection: an original feature and an interaction feature. The original feature is the direct feature of the URL, including special characters in the URL and age of the domain. The interacting feature is the interaction between websites, including in-degree and out-degree of URL.

(ii) We introduce DBN to detect web phishing. We discuss the training process of DBN and get the appropriate parameters to detect web phishing.

(iii) We use real IP flows data from ISP to evaluate the effectiveness of the detection model on DBN. True Positive Rate (TPR) with different parameters is an analysis of our TPR is approximately 90%.

References

1. Detecting Phishing websites using machine learning; Published in: 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)
2. Detection of Phishing Websites from URLs by using Classification Techniques on WEKA; Published in: 2021 6th International Conference on Inventive Computation Technologies (ICICT)
3. Detection and Prevention of Phishing Websites Using Machine; Published in: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) Learning Approach;
4. URL-based Phishing Websites Detection via Machine Learning; Published in: 2021 International Conference on Data Analytics for Business and Industry (ICDABI)
5. https://www.sciencedirect.com/topics/computer-science/phishing-detection