Phishing Web Sites Features Classification Based on Extreme Learning Machine

Yasin Sönmez[1]
Dicle University -Technical Sciences Vocational School
Diyarbakır / Turkey
yasin.sonmez@dicle.edu.tr

Türker Tuncer[2]
Fırat University-Faculty of Technology Forensic Comp.
Elazığ / Turkey
turkertuncer@firat.edu.tr

Hüseyin Gökal [3]
Cyprus International University Faculty of Edu.
Lefkoşa / Cyprus
hgokal@ciu.edu.tr

Engin Avcı[4]
Fırat University-Faculty of Technology Sofware Eng.
Elazığ / Turkey
enginavci@firat.edu.tr

***Abstract*—Phishing are one of the most common and most dangerous attacks among cybercrimes. The aim of these attacks is to steal the information used by individuals and organizations to conduct transactions. Phishing websites contain various hints among their contents and web browser-based information. The purpose of this study is to perform Extreme Learning Machine (ELM) based classification for 30 features including Phishing Websites Data in UC Irvine Machine Learning Repository database. For results assessment, ELM was compared with other machine learning methods such as Support Vector Machine (SVM), Naïve Bayes (NB) and detected to have the highest accuracy of 95.34%**

Keywords—Extreme Learning Machine,Features Classification, Information Security, Phishing.

I. INTRODUCTION

Internet use has become an essential part of our daily activities as a result of rapidly growing technology. Due to this rapid growth of technology and intensive use of digital systems, data security of these systems has gained great importance. The primary objective of maintaining security in information technologies is to ensure that necessary precautions are taken against threats and dangers likely to be faced by users during the use of these technologies [1]. Phishing is defined as imitating reliable websites in order to obtain the proprietary information entered into websites every day for various purposes, such as usernames, passwords and citizenship numbers. Phishing websites contain various hints among their contents and web browser-based information [2-4]. Individual(s) committing the fraud sends the fake website or e-mail information to the target address as if it comes from an organization, bank or any other reliable source that performs reliable transactions. Contents of the website or the e-mail include requests aiming to lure the individuals to enter or update their personal information or to change their passwords as well as links to websites that look like exact copies of the websites of the organizations concerned [6-10].

Phishing Web sites Features

Many articles have been published about how to predict the phishing websites by using artificial intelligence techniques. We examined phishing websites and extracted features of these web sites. Guidelines regarding the extracted features of this database are given below.

In the first section we defined rules and we gave equations of web features. We need these equations in order to explain phishing attacks characaterization.

1.1. Address Bar based Features

1.1.1. Using the IP Address

Rule:

$$\begin{cases} \text{If The Domain Part has an IP Address} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases} \quad (1)$$

1.1.2. Long URL to Hide the Suspicious Part

$$\begin{cases} URL\ length < 54 \rightarrow feature = \text{Legitimate} \\ else\ if\ URL\ length \geq 54\ and\ \leq 75 \rightarrow feature = Suspicious \\ otherwise \rightarrow feature = \text{Phishing} \end{cases} \quad (2)$$

1.3. Using URL Shortening Services “TinyURL”

$$\begin{cases} \text{TinyURL} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases} \quad (3)$$

1.1.4. URL’s having “@” Symbol

$$\begin{cases} \text{TinyURL} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases} \quad (4)$$

1.1.5. Redirecting using “//”

$$\begin{cases} \text{The Position of the Last Occurrence of "//" in the URL} > 7 \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases} \quad (5)$$

1.1.6. Adding Prefix or Suffix Separated by (-) to the Domain

$$\begin{cases} \text{Domain Name Part Includes (−) Symbol} \rightarrow \text{Phishing} \\ \text{Otherwise} \rightarrow \text{Legitimate} \end{cases} \quad (6)$$