



Detecting phishing website using data mining

*Sharmitha S.A**

**First affiliation Chennai, TamilNadu 600107*

ABSTRACT

Now a day's phishing websites are playing major role in internet scams. This project is based on identifying phishing websites using data mining. Phishing websites can be compressed using algorithms. And it can be reported according to other methods data mining gives accurate data of phishing websites. Phishing is social method to steal the personal informations. It can be secured only by identifying the phishing websites and reporting them. Random forest algorithm is used in this project.

Keywords: phishing websites, data mining

1. Introduction

Aim of my project is to create an awareness about phishing websites and how to reporting them. Phishing websites are websites that are created by hackers to appear as a real one. There were at least 68, 688 phishing attacks reported by the Anti-Phishing Working Group in the last five months of 2011. Data mining algorithms used to detect phishing websites in an automated method.

1.1 Objective

This project is to be used to detect phishing websites and not to prevent them. Phishing attack prevention is solution on the user. Our job is just to alert the user whenever we find the URL as phishing. Further methods are completely under users responsibility. Secondly, our project wont cover phishing email detection. Our project only focuses on phishing website detection. it will work only on Chrome. Also, the project will not work on the URL requests made from mobiles.

1.2 Proposed System

A web host model is used for phishing website detection. The model will be based on classification algorithm and will be trained using a training dataset. This model will be deployed online, which will directly communicate with the chrome extension. The detection of the phishing website will be based on URL and website attributes. This system will be a functions carried out on the client and server side. there will be a classifier model trained using the random forest algorithm; a chrome extension will be built and added to the chrome web browser. When the client visits any website using the chrome browser, the URL for the same can be fetched by the chrome extension. From the URL and the displayed webpage, various attributes of the website can be extracted. These extracted URL attributes will act as test data for the classifier deployed in the cloud which can be trained on a phishing website dataset. The classifier will predict the entered URL as either malicious or safe. If it is a phishing website then the user will be alerted that, if they proceed further

on this URL their credentials are at risk of getting misused and if it is a safe website then the user can carry on further operations on that page. Detection and Prevention of Phishing Websites using data mining.

1.3 Module Description

In this module user enter a URL for browse a site, that URL be the input for our machine learning algorithm . Based on the predicted algorithm a user can access the site. If the entered site is phishing site then the user can't access it. module involves five steps. They are 1. Text dataset of Phishing site details 2. In pre-processing step, separating data and label 3. In feature extraction stage, the text features are extracted. The system cannot understand the text data so we convert the text data into numerical values.

1.4 Significance

This system can be used by so much E-commerce or other websites in order to have good customer relationship. User can make online payment . Data mining algorithm used in this system provides better performance as compared to other traditional classification algorithms. With the help of this system user can also purchase products online without any hesitation.

1.5 Methodology:

Detection and Prevention of Phishing Websites using Data mining Approach is a mechanism that is proposed in order to ensure high security for data. In the proposed system, a cloud-based model is used for phishing website detection. The model will be based on classification algorithm and will be trained using a training dataset method. This model will be deployed in the cloud, which will directly communicate with the google chrome extension. The detection of the phishing website will be based on URL and website servers..

1.6 Originality of the Project

Detecting phishing websites using data mining is to create a awareness about phishing websites and to have secure browsing experience with any data or personal information loss. And also to prevent the credentials of the users from the attackers .this projects will find the phishing websites with accuracy of 90% result.

1.7 Conclusion

Through the system, the goal is to implement the detection of the phishing websites using data mining. This work will be done by extracting the features of the website URL when the user go through it. The collected url will act as test data for the model. In order to detect the phishing website, we proposed an intelligent, and effective system that is based on using the new Data mining algorithm. We implemented classification algorithm and techniques to extract the phishing data sets criteria to analyze their legitimacy to stop the attackers.

REFERENCES

-
- a. Ahmed, Abdulghani Ali, and Nurul Amirah Abdullah. "Real time detection of phishing websites." Information Technology, Electronics and Mobile Communication Conference (IEMCON), 7th Annual. IEEE, 2016.
 - b. Singh, Priyanka, Yogendra PS Maravi, and Sanjeev Sharma. "Phishing websites detection through supervised learning networks." Computing and Communications Technologies (ICCCT), 2015 International Conference on. IEEE, 2015
 - c. Jain, Ankit Kumar, and B. B. Gupta. "Comparative analysis of features based machine learning approaches for phishing detection." Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016