MINOR PROJECT 2

MID SEM REPORT
on

Intelligence Rule-Based Phising Websites Classification Based On URL Features

Submitted By :

Name	Roll No	Branch
Abhinav Singh	R970216002	CSE OGI
Aniket Shankar	R970216012	CSE OGI
Ankit Jha	R970216013	CSE OGI
Anubhav Raj	R970216016	CSE OGI

Under the guidance of

Mr. Nitin Arora

Assistant Professor

Department of Informatics

School of Computer Science

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

Dehradun-248007

2019-20

Approved By

(Mr. Nitin Arora)
Project Guide

(Dr. Thipendra Pal Singh)
Department Head

ABSTRACT

Now a days phising attack has turned out to the most serious issues faced by internet users, organizations and service providers. In phishing attack attacker tries to obtain the individual data of the users by using spoofed emails or by using fake websites. The internet community is still looking for the complete solution to secure the internet from such attacks. As the trend of internet is increasing the cyber attacks are also increased. Indian enterprise face over 2.8 lakhs of cyber attacks daily. In order to stop this cyber attacks we are trying to classify the websites as phishing website and the original websites. In this project we create a extension software to classify the website as Phishing Website or not. It is used to detect phishing website and perform various types of studies.

Keywords: Phishing website, Anti-Phishing, Input form, Phishers, Extension

TABLE OF CONTENTS

Contents

LIST OF FIGURES

List of Figures

1 Introduction

Phishing, as the act of stealing personal information of Internet users for misuse is an old but still threatening problem. It is a form of identity theft that occurs when a malicious website impersonates a legitimate one in order to acquire confidential information such as account details, user passwords, etc. Though there are several anti-phishing softwares and techniques for detecting potential phishing attempts in emails and detecting phishing information on web pages, Phishers come up with new and reduction techniques to circumvent the available software and techniques. The phishing attackers mislead users by employing different social engineering tactics such as threatening to suspend user accounts if they do not complete the account update process, furnish other information to validate their accounts or some reasons to get the users to visit their spoofed web sites. Phishing attacks affect millions of internet users and are a huge cost burden for businesses and victims of phishing. Phishing has become a serious threat to users and businesses alike. Over the past few years, plenty of attention has been paid to the issue of security and confidentiality. As the amount of Internet users and online transactions multiples, the possibility of misuse is also growing. Phishing is hence an important cyber security issue.

2 Background Study

Phishing website is a huge effect on the financial and online transactions, detecting and preventing this attack is an important step towards protecting against website phishing attacks, there are many approaches to detect these attacks. In this section, we review about existing anti phishing solutions and list of the related study.

- One approach is, client-side defense against web-based identity theft . It proposes a framework for client-side defense: a browser plug-in called Spoof Guard that examines web pages and warns the user when requests for data may be part of a spoof attack, it calculates a spoof index (a measure of the likelihood that a specific page is part of a spoof attack), and alert the user if the index exceeds a level selected by the user. Spoof Guard uses both combination of outgoing post data examination to compute a spoof index and page evaluation. When a user enters a username and password on a spoof website that contains some combination of suspicious misleading domain name, URL, images from an honest website, and password and a username that have previously been used at an honest website, Spoof Guard will block the post and warn the user with a popup that foils the attack.[1]

- In the paper of O.Kalaiselvan it describe the phishing and its characteristics e URL is long, IP address in URL, join prefix and suffix to domain and request URL and etc. It basically focuses on the anti phishing softwares and techniques. The attackers basically mislead the user by employing different engineering tactics such as threghten to suspend user accounts. As the amount of internet user and online banking is increasing exponentially so the phishing issue is also increasing at a very rapid rate. Now days the Large number of phishers used the sophiscated software kit to launch a large number of phishing website on the different URL to the common security methods. Issue that the phishers usually try to closely by impersonate a trusted party the user knows by imitating brands,web design,logo or a special case the URL. Classification is one of the most important technique for the data mining.Through this we could classify various fishing website based on some features and take the necessary actions.[2]

- In the paper of Rami M. Mohammad it describe the phishing as a art of emulating the website and try to grab the user private information such as the username, password security number etc. The attacker basically send an email and ask an individual to update their personal information by clicking the link so that the attacker can get their personal information. Now days various feature had been included so that this could help to diffrentiate from the phishing website.Their are two approaches to identify the phishing website first is Blacklist and the second is the heuristic-based methods.In this paper it differ from all other researcher by including a group of feature,these feature are examine in predicting phishing website by using rule-induction algorithm that focus on to reduce the negative rate.Automatic extraction is faster than the manul extraction.An experiment was conducted that showed that the we could improve the predicting accuracy on some features like URL,Age of Domain, HTTPS and SSL, Website Traffic and etc.[3]

3 Problem Statement

Phishing web pages are fake websites generated by dishonest people to impersonate original web page.Users may not be able to access their emails or sometimes lose money because of phishing. Predicting and blocking this attack is a critical step toward protecting online transactions. The efficiency of predicting the type of the website necessarily depends on the extracted features goodness. Since most of the users feel safe against phishing attacks if they utilize an anti phishing tool, this deliver a great responsibility on the anti phishing tools to be accurate in predicting phishing.

4 Objective

To classify phishing websites based on their URL features,which can be used to detect these type of websites in future using various Machine learning classification algorithm.Here we create a extension software to classify the website as Phishing Website or not. It is used to detect phishing website and perform various types of studies.

5 Methodology

- Gathering of information about the topic from various sources.
- Selection of problem from the information gathered.
- Design a chrome Extension
- Deployment of extension in chrome
- Selection of classification algorithm
- Implementation of classification algorithm
- Train the Extension through dataset
- Testing and displaying the output

6 Design

6.1 Use Case Diagram

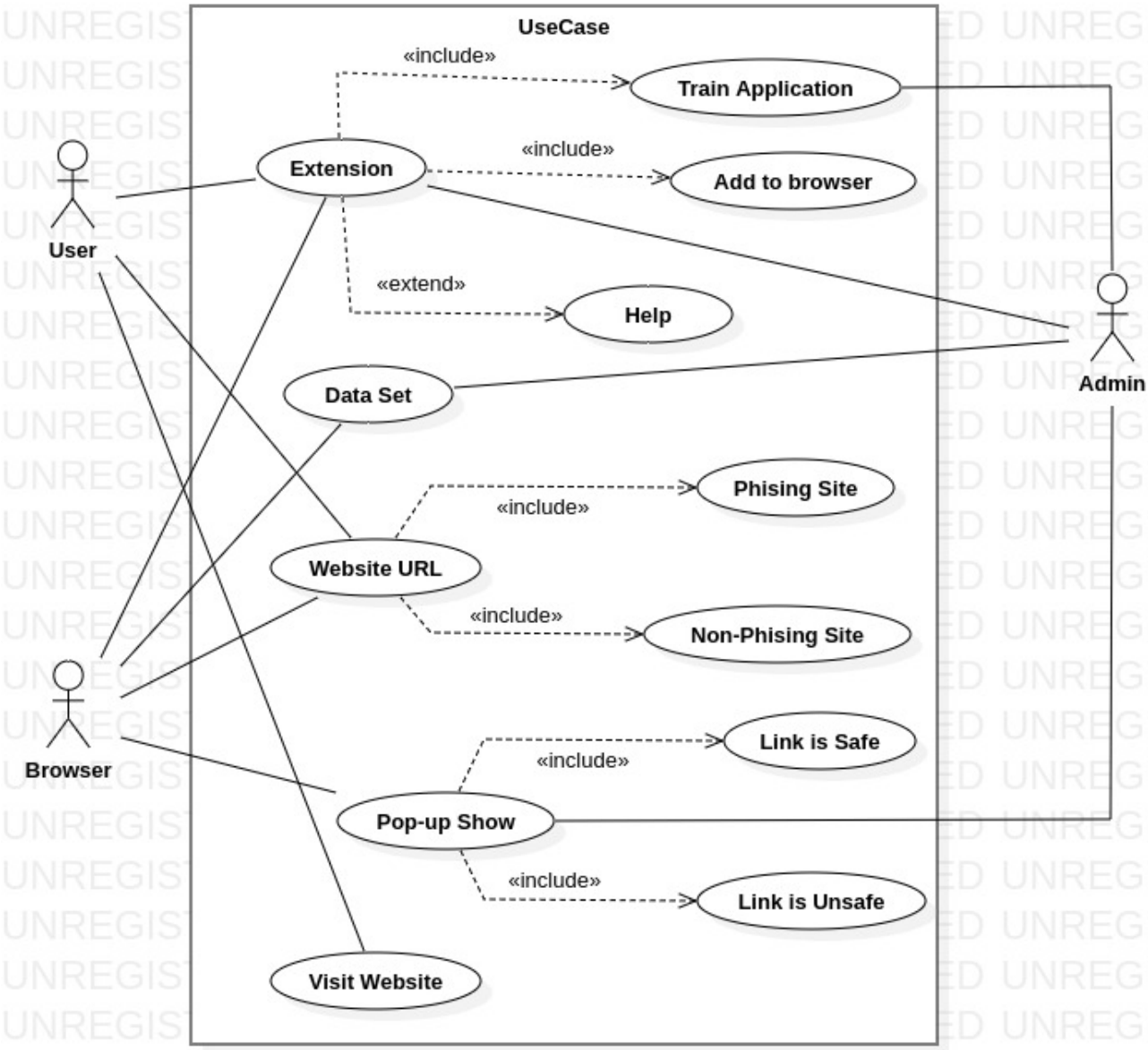


Figure 1: Use Case Diagram

6.2 Activity Diagram

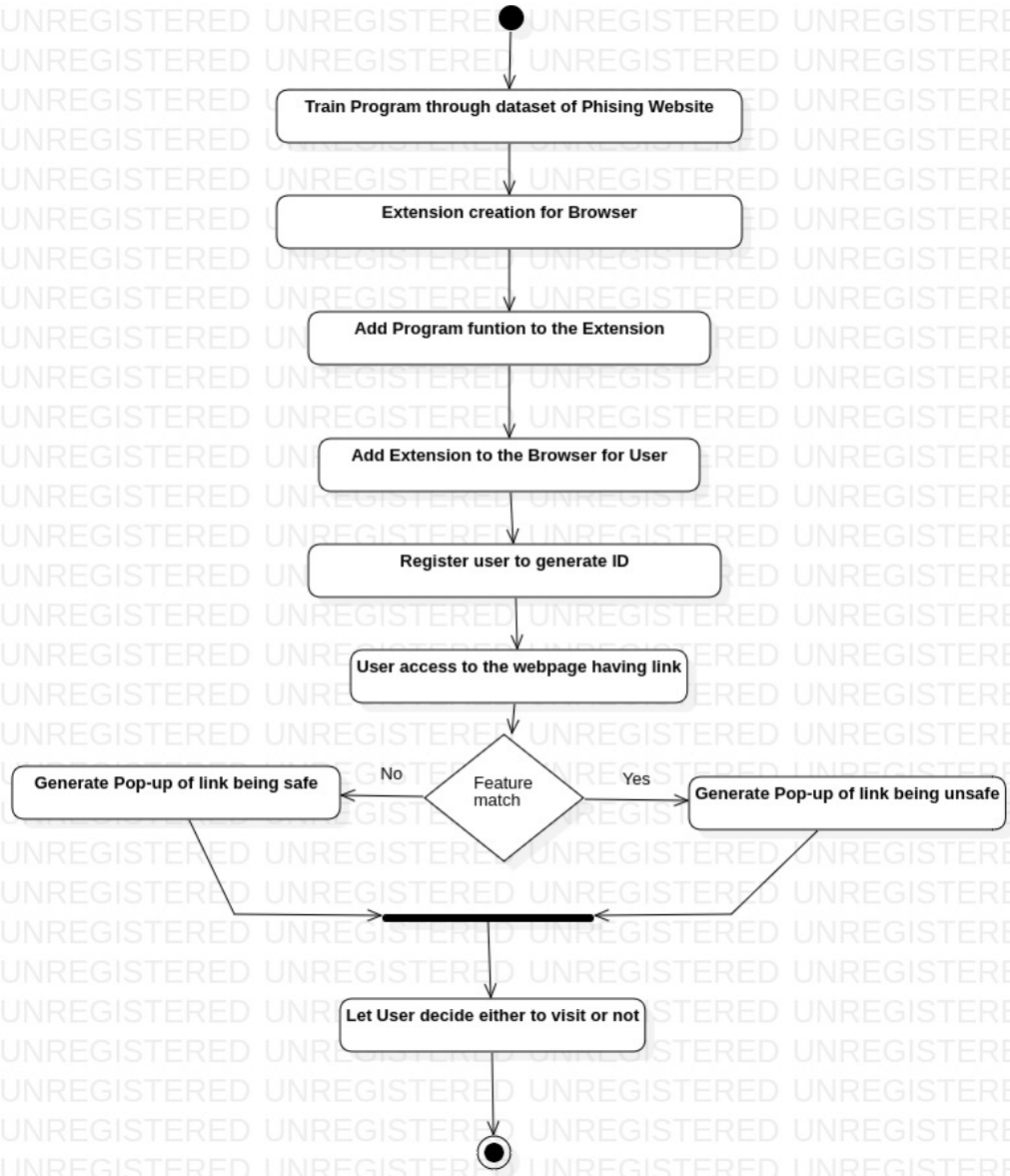
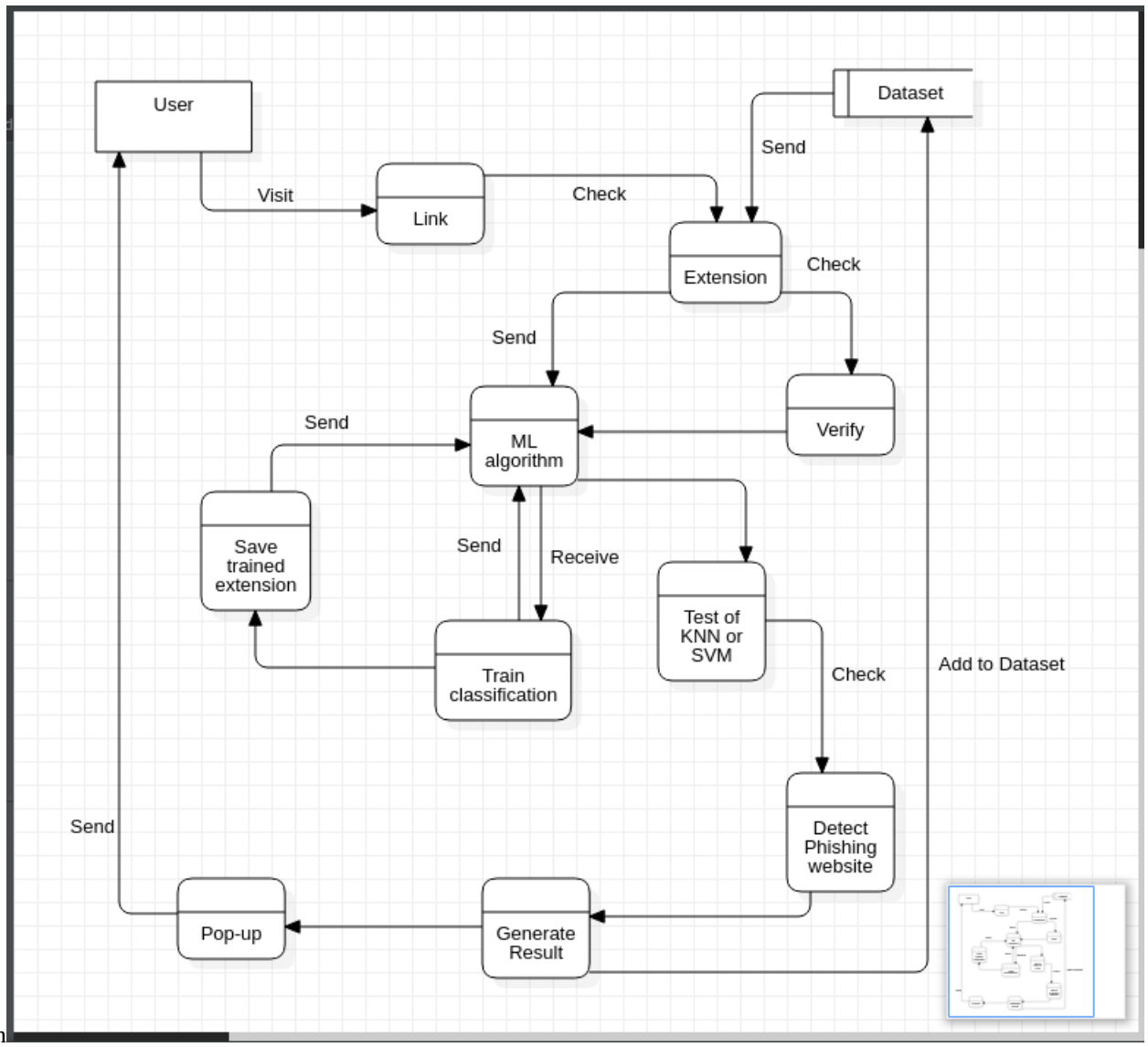


Figure 2: Activity Diagram

6.3 Data Flow Diagram



widthwidth

Figure 3: Data Flow Diagram

7 Implementation

Create a folder in which we will store all the files of our extension.

- Create a JSON file named manifest.json.
- Store all the information about extension in this file.
- This JSON file included- name of extension,version of extension,manifest version used, description of extension, about the action that the extension will take, which itself include the icon to be used for extension, and the HTML file used to build extension, and permission that where the extension can be used, in our case all websites.
- Then we will build an HTML file which will contain the main content of the extension and will call the javascript file for external functionality to be added
- There are two javascript file one named popup.js and getPagesSource.js.
- popup.js is calling getPagesSource.js to perform its functionality as well as to print error message if some network issue or any thing else arises.
- Whereas,getPagesSource.js file include the code to extract URLs from the current page.

7.1 Pseudo Code

popup.html

- declare DOCTYPE html
- initiate html tag with style value
- initiate head tag
- initiate script tag with src value as 'popup.js'
- end script tag
- end head tag
- initiate body tag with style="width:500px"
- initiate div tag with id='message'
- print 'Extracting URLs....' message
- end div tag
- end body tag
- end html tag

popup.js

- declare chrome.runtime.onMessage.addListener registration with function having arguments request and sender
- if request.action equals "getSource"
- assign request.source to message.innerText
- declare function onWindowLoad define variable message and assign value document.querySelector('mess

- declare chrome.tabs.executeScript with parameters as null, assign file as ”getPagesSource.js”
- declare function if thier is chrome runtime error assign ’There was an error injecting script : n’ + chrome.runtime.lastError.message to message.innerText
- call function onWindowLoad from ”getPagesSources.js” file on window load

getPagesSource.js

- declare function urlDetect with parameter as documentroot
- declare variable links and assign document.getElementsByTagName(’a’)
- call alert(links.length) to print no. of URLs in links variable
- for variable i = 0 to links.length
- print in console links[i].href
- declare chrome.runtime.sendMessage registration
- action as ”getSource”,
- source as calling urlDetect function with argument document

References

[1] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, J. C. Mitchell, and S. Ca, "Client-side defense against web-based identity theft,"

[2] S. Edwinraja and P. S. R. E. College, "Predicting Phishing Websites using Rule Based T ECHNIQUES," vol. I, no. 4, pp. 180–185, 2015.

[3] R. M. Mohammad, F. Thabtah, and L. Mccluskey, "Intelligent Rule based Phishing Websites Classification," pp. 1–22.