

Heuristic Based Approach for Phishing Site Detection

Project Guide:- Prof. Sunil A. Sushir

GROUP MEMBERS:-

1]CHIRAG CHAUDHARI (15204030)

2] SWAPNIL KSHETRE (15204003)

3]SWAPNIL GHAWALI (15204009)

4] AAKASH SANE (15204012)

Index

- Abstract
- Introduction
- Literature Review
- Problem Statement
- Project Scope
- Use Case Diagram
- Architecture
- Project Future Plan
- Summary
- Prototype
- Referance

Abstract

- Phishing has been a major security threat in which there is a huge loss for companies as well as customers.
- These phishing attacks are increasing day by day due to lack of efficient detection techniques and effective preventive measures.

Contd...

- This paper proposes a Heuristic-based phishing detection algorithm. In particular, this research focuses on improving upon the previously published text-based approach.
- The algorithm in the previous work analyzes the body text in an email to detect whether the email message asks the user to do some action such as clicking on the link that directs the user to a fraudulent website

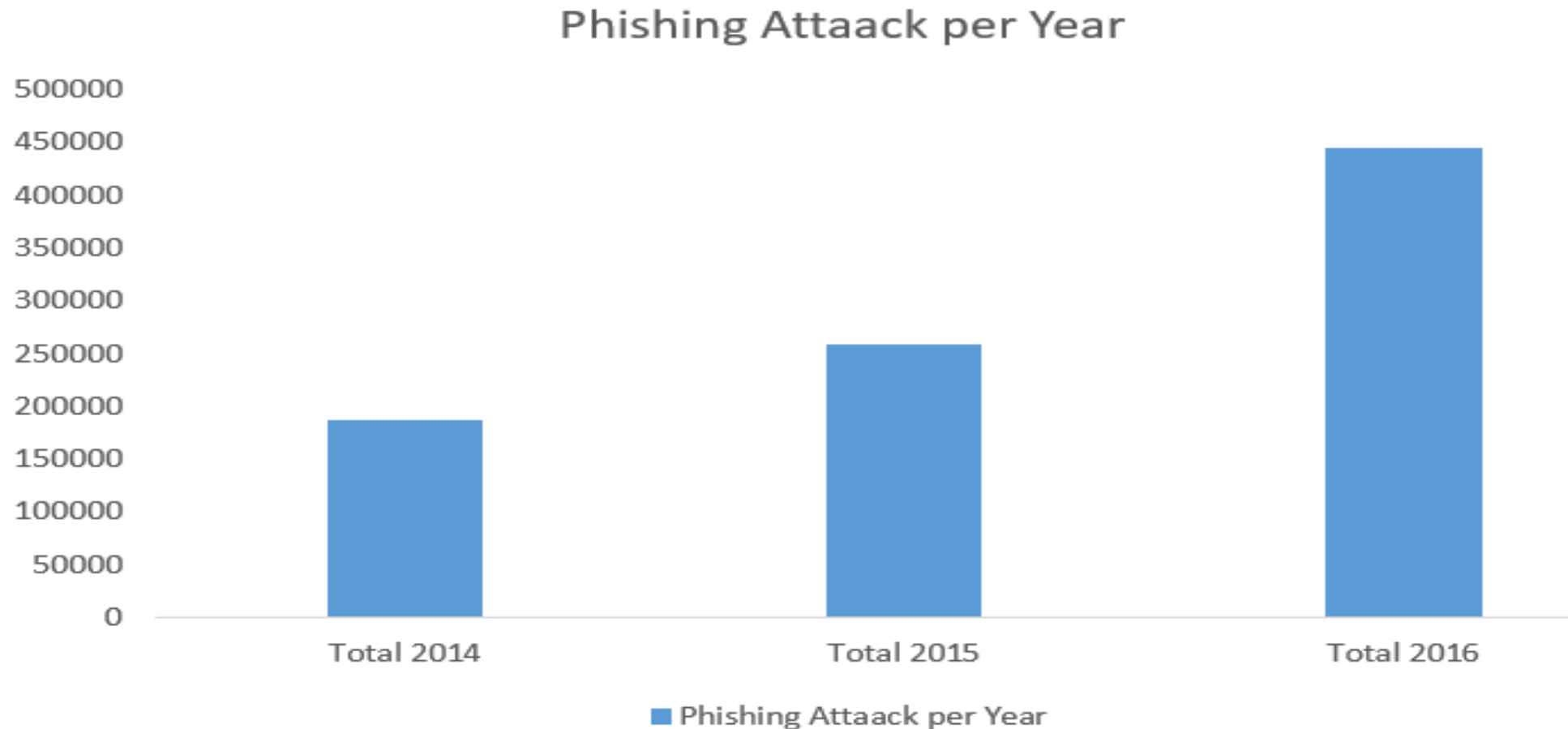
Introduction

- Phishing is a type of social engineering attack that targets a user's sensitive information through a phony website that appears similar to a legitimate site, or by sending a phishing email [1].
- Phishing is a malicious use of Internet resources carried out to trick Internet users to reveal personal information, such as usernames, credit card information, and Social Security numbers to the attacker[2].
- Various approaches have been proposed and implemented to detect a variety of phishing attacks such as use of blacklists and whitelists

Contd...

- We propose a desktop application called PhishSaver, which focuses on URL and website content of the phishing web page.
- We aim at detecting phishing websites using Decision tree - Machine learning technique with the help of a desktop application named PhishSaver.
- Phish-Saver use a heuristic features to detect a number of phishing attacks

Phishing Attack Survey



Literature Review

Title	Problem Identified	Methodology	Strength	weakness
The Sceured Anti-Phishing Approach Using image based Validation. Y. Yesu Jyothi, D.Srinivas & k.GovindaRaju,2013	To solve the problem of phishing & protect individual personal private information	Visual Cryptography (image based validation)	It prevents attack of phishing websites on financial portal,banking portal & online shopping market	Inability to recover missing or corrupt share
Protecting users Against phishing attacks. Engin kirda & Chistopher Kruegel,2012	Increased email linked to phishing scams	Browser Extension	It protect users against spoofed website – based phishing attacks	It requires that user support to capture & store sensitive information rather than automatically captureing & storing the sensitive information
Phishnet Anti-Phishing Technique(Prakash et al.,2010)	Predicts variation of URLs	Heuristics	It replace Top level domain(TLD),Dirctory structure similarity,IP address equivalence,Qury string substitution & brand name equivalence	It cannot detect zero day phishing

Problem Statement

- Phishing has been a major security threat in which there is a huge loss for companies as well as customers. These phishing attacks are increasing day by day due to lack of efficient detection techniques and effective preventive measures.
- A comprehensive efficient detection technique should be developed in order to detect and inform the web users about the phishing attacks to make sure that their sensitive data will not be disclosed during these attacks.
- This research project deals with a comprehensive heuristic based method for phishing detection which is based on content of the website through which phishing attacks can be discovered

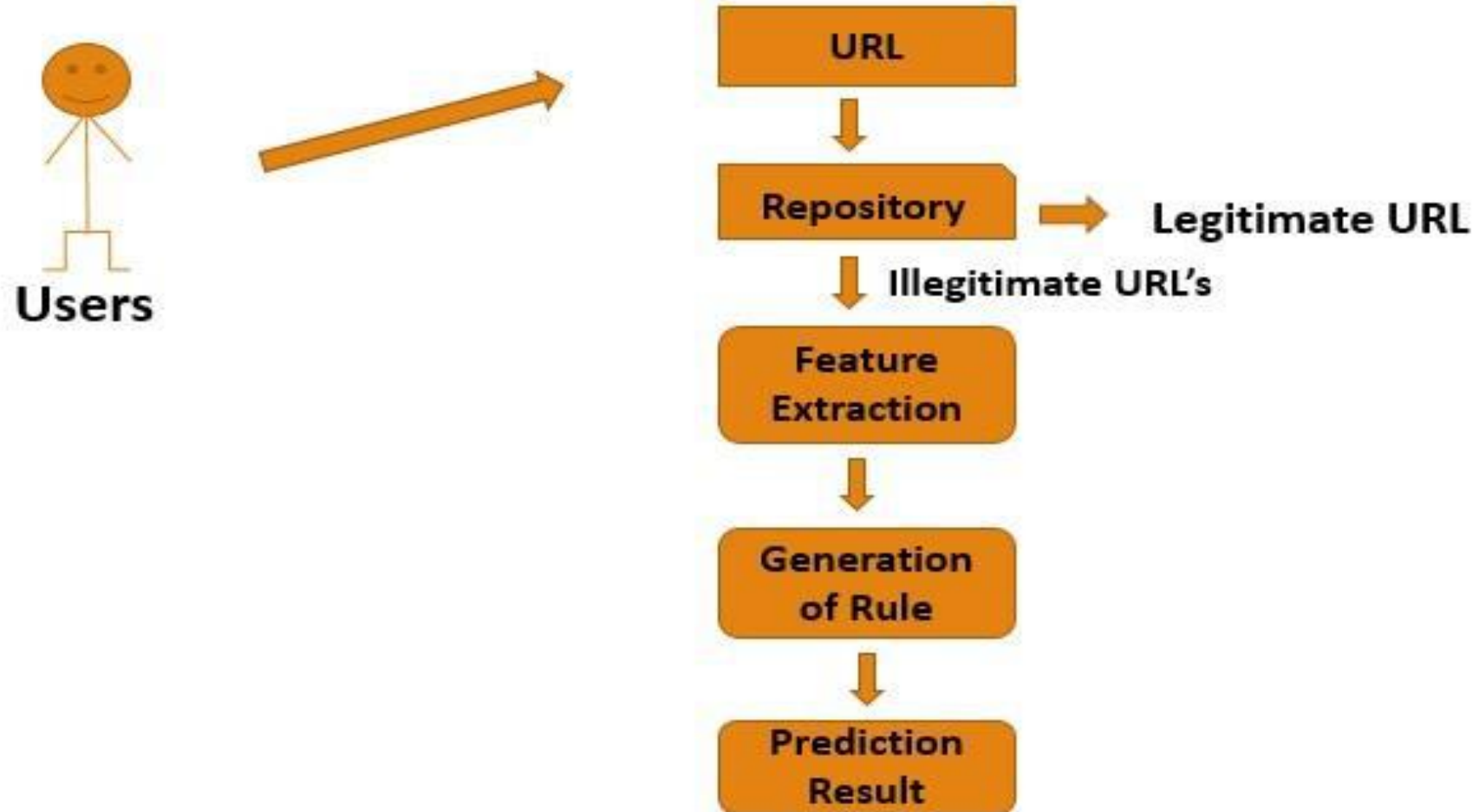
Project Scope

- In this project URL- based heuristic approach will be used along with the ranking of sites to extract the features from the URL.
- All the extracted features along with the phishing and legitimate sites URLs will be stored in database.
- A classifier will be generated using decision tree algorithm which will classify the URLs as phishing and legitimate.

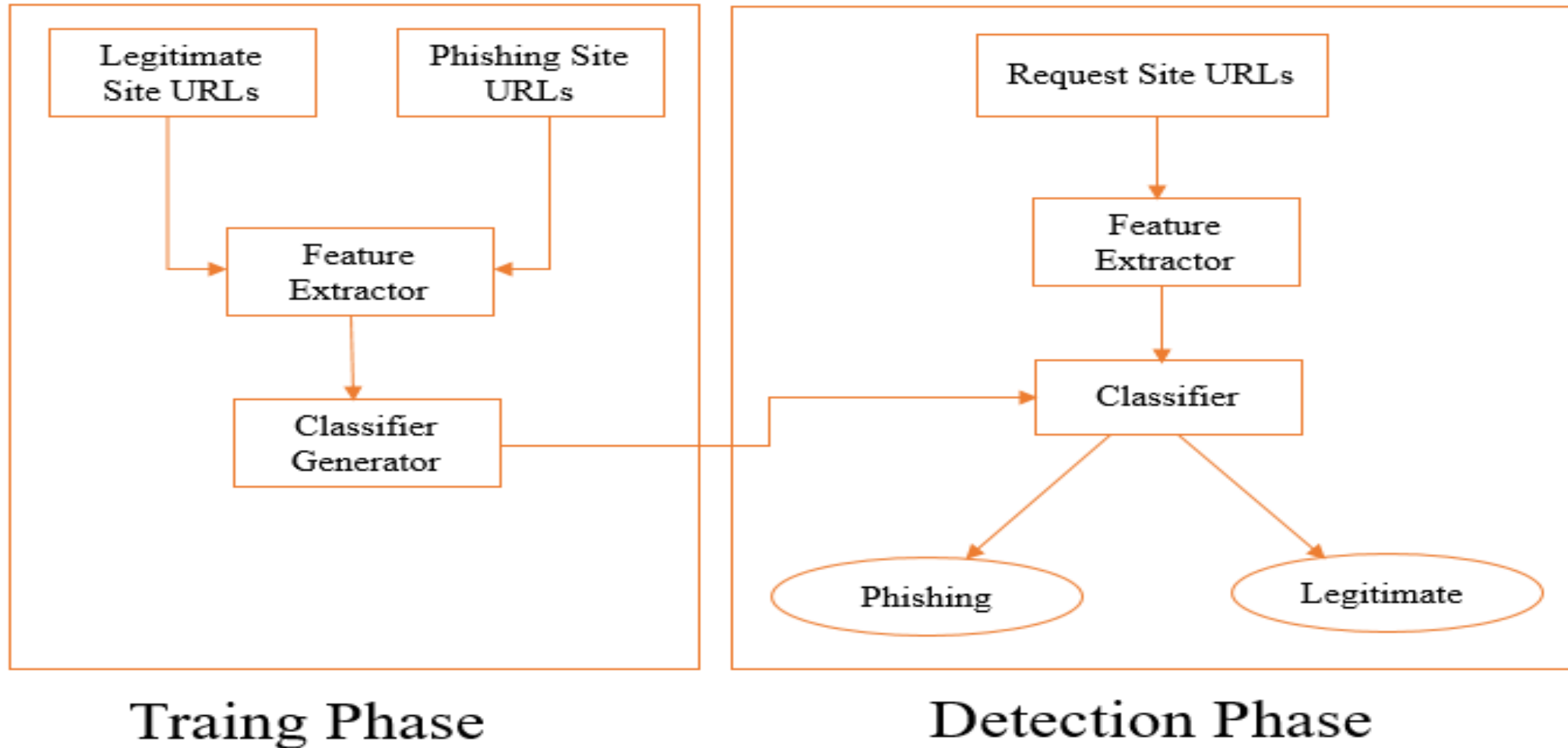
Contd...

- When new URL is received it will extract the features and will compare them with the features stored in database, thus classifying the incoming site as phishing or legitimate.
- This rule induction helps to facilitate the decision making process which ensures reliability and completeness

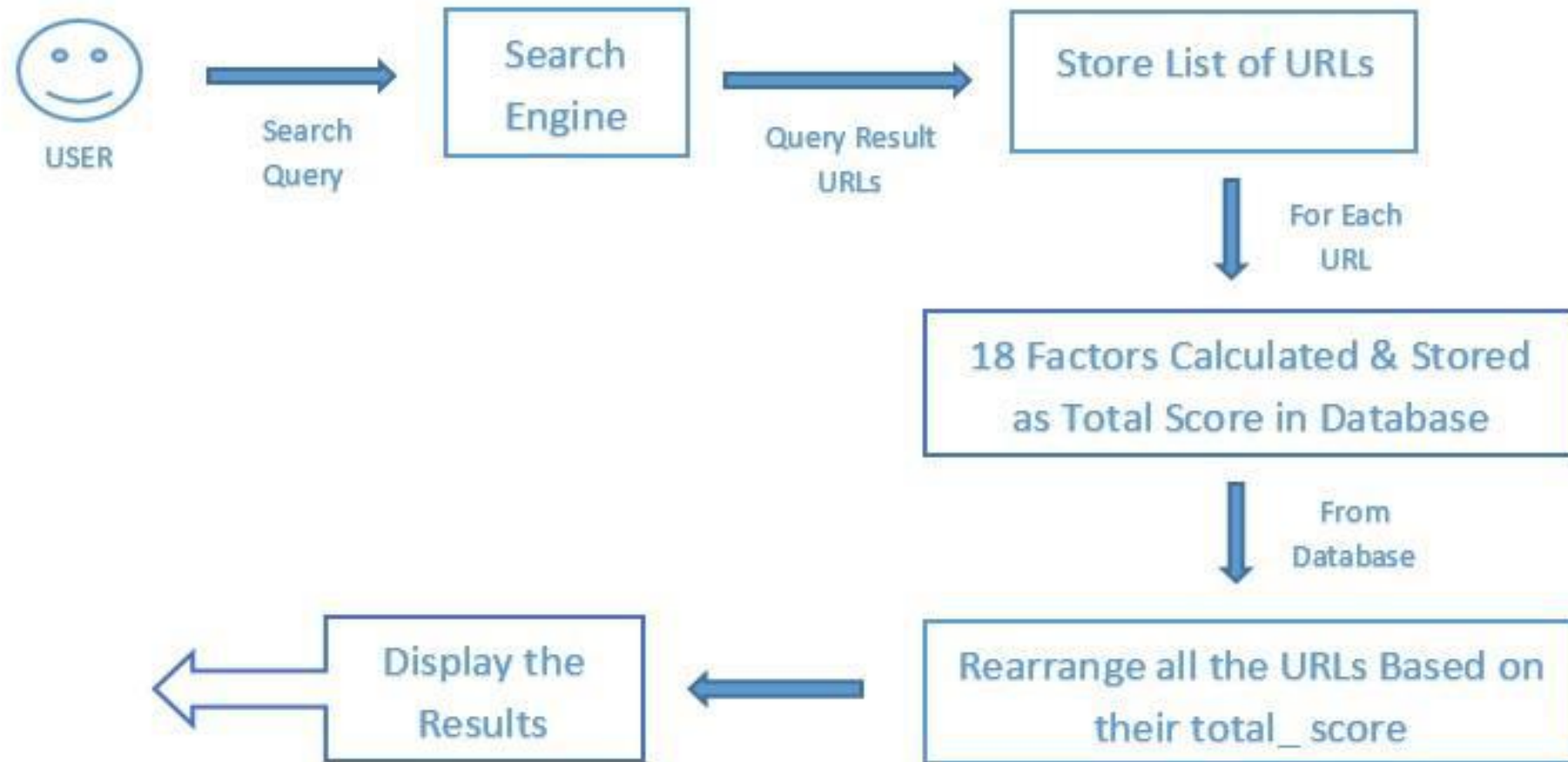
Use case Diagram



Architecture



Block Diagram



Project Future Plan

- In future work, we intend to address the time-intensive disadvantage of the heuristic-based technique. With a large number of features, it is time-consuming for the heuristic-based approach to generate classifiers and perform classification . Therefore, we will apply algorithms to reduce the number of features and thereby improve performance.
- In addition, we will examine a new phishing detection technique that uses not only URL-based features, but also HTML and JavaScript features of web pages to improve performance.
- Also make plugin for the browser which will alert user about phishing website and reduce damage cause with it as much as possible.

Summary

- The proposed model can reduce damage caused by phishing attacks because it can detect new and temporary phishing sites.
- System also implemented decision tree algorithm and generated tree for it We will be looking forward for the new features to use and try to improve more accuracy and reduce false positive value of the system.
- We also look forward to discover new feature with high impact to detect phishing.

Prototype

The image shows a prototype of a web application window. The window has a title bar with the text "CLASSIFY-PHISH" and standard window control buttons (minimize, maximize, close). Below the title bar is a menu bar with the word "File". The main content area has a light blue background. It contains a label "ENTER URL:" followed by a white text input field. Below the input field are two buttons: "CHECK" and "CLEAR". The "CHECK" button is light gray, and the "CLEAR" button is light blue with a blue border.

File

ENTER URL:

<http://rdacodisha.org/secured/>

CHECK

CLEAR



CLASSIFY-PHISH



File

ENTER URL:

http://rdacodisha.org/secured/

CHECK

CLEAR



Enter Valid URI

OK

LENGTH :

HTTP Present :

SUSPICIOUS CHARACTER :

NUMBER OF DOTS :

LENGTH OF SUB DOMAIN :

NUMBER OF "/" IN URL :

File

ENTER URL:

CHECK

CLEAR

File

ENTER URL: **LENGTH :** 21**HTTP Present:** yes**SUSPICIOUS CHARACTER :** suspicious character not found**NO. OF DOTS :** 2**LENGTH OF SUB DOMAIN :** 3**NO OF "/" IN URL :** 2

References

- [1] Khonji, Mahmoud, Youssef Iraqi, and Andrew Jones . "Phishing detection: a literature survey. "Communications Surveys Tutorials ,IEEE 15.4(2013): 2091-2121.
- [2] APWG, Phishing activity trends paper.[online]. [http://docs.apwg.org/reports/APWG](http://docs.apwg.org/reports/APWG%20Global%20Phishing%20Report%201H%202014.pdf) Global Phishing Report 1H 2014.pdf
- [3] Luong Anh Tuan Nguyen¹, Ba Lam To¹, Huu Khuong Nguyen¹ and Minh Hoang Ngu-yen²¹ Faculty of Information Technology, 2014 IEEE An Efficient Approach for Phishing Detection Using Single-Layer Neural Network
- [4] So Young Rieh, Judgment of Information Quality and Cognitive Authority in the Web, citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.107.8991
- [5] Suman Bhattacharyya, Chetan Kumar Pal, Praveen Kumar Pandey, Detecting Phishing Websites, a Heuristic Approach, International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137 Volume 02, Issue 03, March 2017, PP 120-129

Contd...

[6] Nguyen, Luong Anh Tuan, et al. "A novel approach for phishing detection using URL-based heuristic." Computing, Management and Telecommunications (ComManTel), 2014 International Conference on. IEEE, 2014

[7] Wikipedia.(2015.March) Uniform Resource Locator . Available:
[http://en.wikipedia.org/wiki/Uniform resource locator](http://en.wikipedia.org/wiki/Uniform_resource_locator)

[8] Sunil, A. Naga Venkata, and Anjali Sardana. "A pagerank based detection technique for phishing web sites." Computers Informatics (ISCI), 2012 IEEE Symposium on. IEEE, 2012.

[9] WANG, Wei-Hong, et al. "A Static Malicious Javascript Detection Using SVM." strings. Vol. 40. 2013

[10] Hou, Yung-Tsung, et al. "Malicious web content detection by machine learning." Expert Systems with Applications 37.1 (2010): 55-60.



Thank You!