Basic Details of the Team and Problem Statement

PSID (Problem statement ID) : **KVH-004**

Problem Statement Title: **Phishing Detection Solution**

Team Name: Cyber Knights

Team Leader Name: Santhosh Raminedi

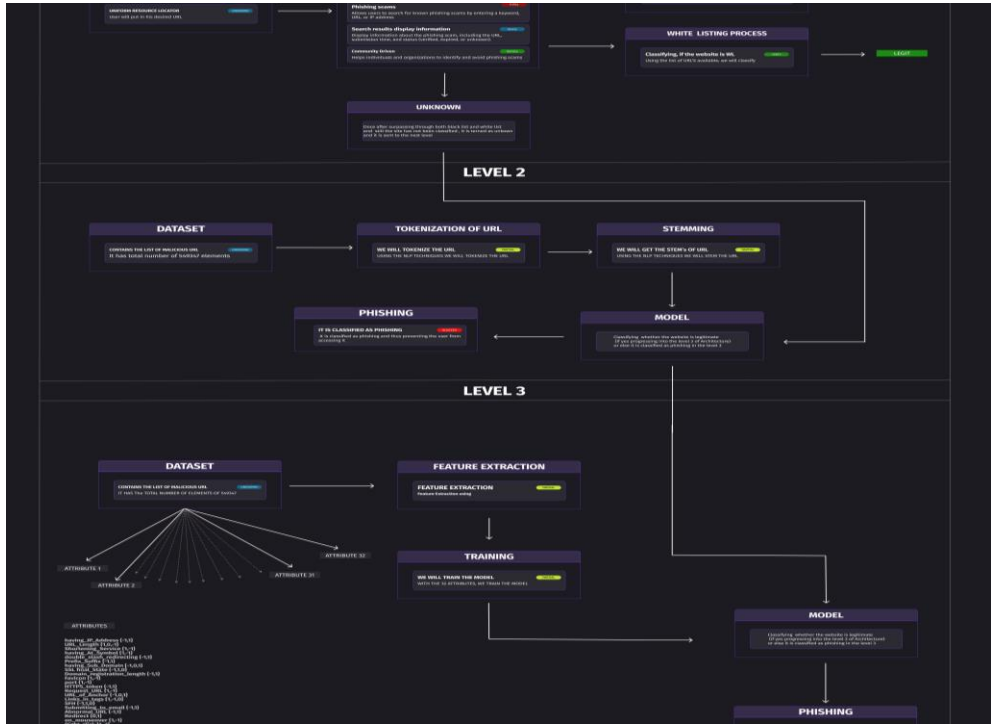
Institute Code (AISHE):C-49441

Institute Name: Vellore Institute of Technology, Chennai

Idea/Approach Details

Describe your idea/Solution/Prototype here:

- Proposed a 3-level Phishing architecture.
- Level -1 : Blacklisting and Whitelisting module
- In which phishing URL's and Legitimate URL's are extracted from the Phish Tank using web scrapping methods (1.8 lakh Phishing and 60k Legitimate). Given URL is compared with these links as a preliminary step.
- Level – 2: Visual Segmentation module
- In this module the URLs are tokenized and extracted stem of the tokenized words. From this keywords we have performed machine learning algorithms to classify phishing and legitimate website URLs.
- Level – 3 URL Behavior Observation module
- Features of the URL are extracted using web mining tools. Features extracted as categorized as address, domain and HTML features of the website URL. Based on this features of the URLs we will make a classification using Machine Learning algorithms.



Describe your Technology stack here:

- Machine Learning (Logistic Regression, Gaussian Naïve Bayes,
- Flask.
- Manifest.
- Python.
- Web mining

Idea/Approach Details

Describe your Use Cases here (a specific situation in which a product or service could potentially be used)

- 1. Email Phishing detection: Users are becoming prey to the deceptive tactics of the phishers. Web plugin will help in indicating the whether browsing URLs have phishing or legitimate.
- 2. Similar website URLs: In this case phishing attackers use similar kind of original URL but small variation in the character of the URL. Users may not find and lose sensitive information. 2nd module of the idea is efficient in classifying this type of the website phishing attacks.
- 3. One day attack: Phishing softwares use 3rd parties to get the phishing links (which is blacklisting and whitelisting methods). But the newly created website url may not be included in the list of links. For this attacks 3rd module which extracts the meta information of the URL. Age of the domain plays an important role in classifying this type of websites.

Describe your Dependencies / Show stopper here

(*Dependency means something without which your project cannot be developed…could be hardware or software…

*Show stoppers are threats that can arise and can delay the development of the solution…)

- Dependency for the module 1: Phish Tank (from which we extract URLs to classify in the preliminary step)
- Dependency for the module 2 and module 3: Dataset of the phishing and Legitimate website URLs.
- Web Plugin (It is important to extract the URLs of the active tab)
- Flask server (Which is essential for running the server and loading the trained models)
- No of URLs in the active tab vs Response time

Team Member Details

Sr. No.	Name of Team Member	Branch (Btech/Mtech/PhD etc):	Stream (ECE, CSE etc):	Year	Position in team (Team Leader, Front end Developer, Back end Developer, Full Stack, Data base management etc.)
1	Santhosh Raminedi	20BCE1477	CSE	2020	Team Leader
2	Rahul Sandireddy	20BCE1001	CSE	2020	
3	Venkat Amith Woonna	20BCE1222	CSE	2020	
4	Krishna Prasad Y V S Purama	20BCE1421	CSE	2020	
5	Vemasani Varshini	20BCE1240	CSE	2020	
6	Amogh Singh	20BRS1149	CSE	2020	

Team Mentor/s Details (Mandatory)

Sr. No.	Name of Mentor	Category (Academic/Industry):	Expertise (AI/ML/Blockchain etc):	Domain Experience (in Years)
1	Menaka Pushpha A	Academic	AI/ML and Network security	10
2				