WEB PHISHING DETECTION

Problem Statement

There are a number of users who purchase products online and make payments through e-banking. There are e-banking websites that ask users to provide sensitive data such as username, password & credit card details, etc., often for malicious reasons. This type of e-banking website is known as a phishing website. Web service is one of the key communications software services for the Internet. Web phishing is one of many security threats to web services on the Internet.

Common threats of web phishing:

- Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity.

- It will lead to information disclosure and property damage.

- Large organizations may get trapped in different kinds of scams.

This project mainly focuses on applying a machine-learning algorithm to detect Phishing websites.

Literature Survey On The Selected Project & Information Gathering

In this activity you are expected to gather/collect the relevant information on project use case, refer the existing solutions, technical papers, research publications etc.

H. Huang et al., (2009) proposed the frameworks that distinguish the phishing utilizing page section similitude that breaks down universal resource locator tokens to create forecast preciseness phishing pages normally keep its CSS vogue like their objective pages.

S. Marchal et al., (2017) proposed this technique to differentiate Phishing website depends on the examination of authentic site server log knowledge. An application

Off-the- Hook application or identification of phishing website. Free, displays a couple of outstanding properties together with high preciseness, whole autonomy, and nice language-freedom, speed of selection, flexibility to dynamic phish and flexibility to advancement in phishing ways.

Mustafa Aydin et al. proposed a classification algorithm for phishing website detection by extracting websites' URL features and analyzing subset based feature selection methods. It implements feature extraction and selection methods for the detection of phishing websites. The extracted features about the URL of the pages and composed feature matrix are categorized into five different analyses as Alpha-numeric Character Analysis, Keyword Analysis, Security Analysis, Domain Identity Analysis and Rank Based Analysis. Most of these features are the textual properties of the URL itself and others based on third parties services.

Samuel Marchal et al. presents PhishStorm, an automated phishing detection system that can analyze in real time any URL in order to identify potential phishing sites. Phish storm is proposed as an automated real-time URL phishingness rating system to protect users against phishing content. PhishStorm provides phishingness score for URL and can act as a Website reputation rating system.

Fadi Thabtah et al. experimentally compared large numbers of ML techniques on real phishing datasets and with respect to different metrics. The purpose of the comparison is to reveal the advantages and disadvantages of ML predictive models and to show their actual performance when it comes to phishing attacks. The experimental results show that Covering approach models are more appropriate as anti- phishing solutions.

Muhemmet Baykara et al. proposed an application which is known as Anti Phishing Simulator, it gives information about the detection problem of phishing and how to detect phishing emails. Spam emails are added to the database by Bayesian algorithm.

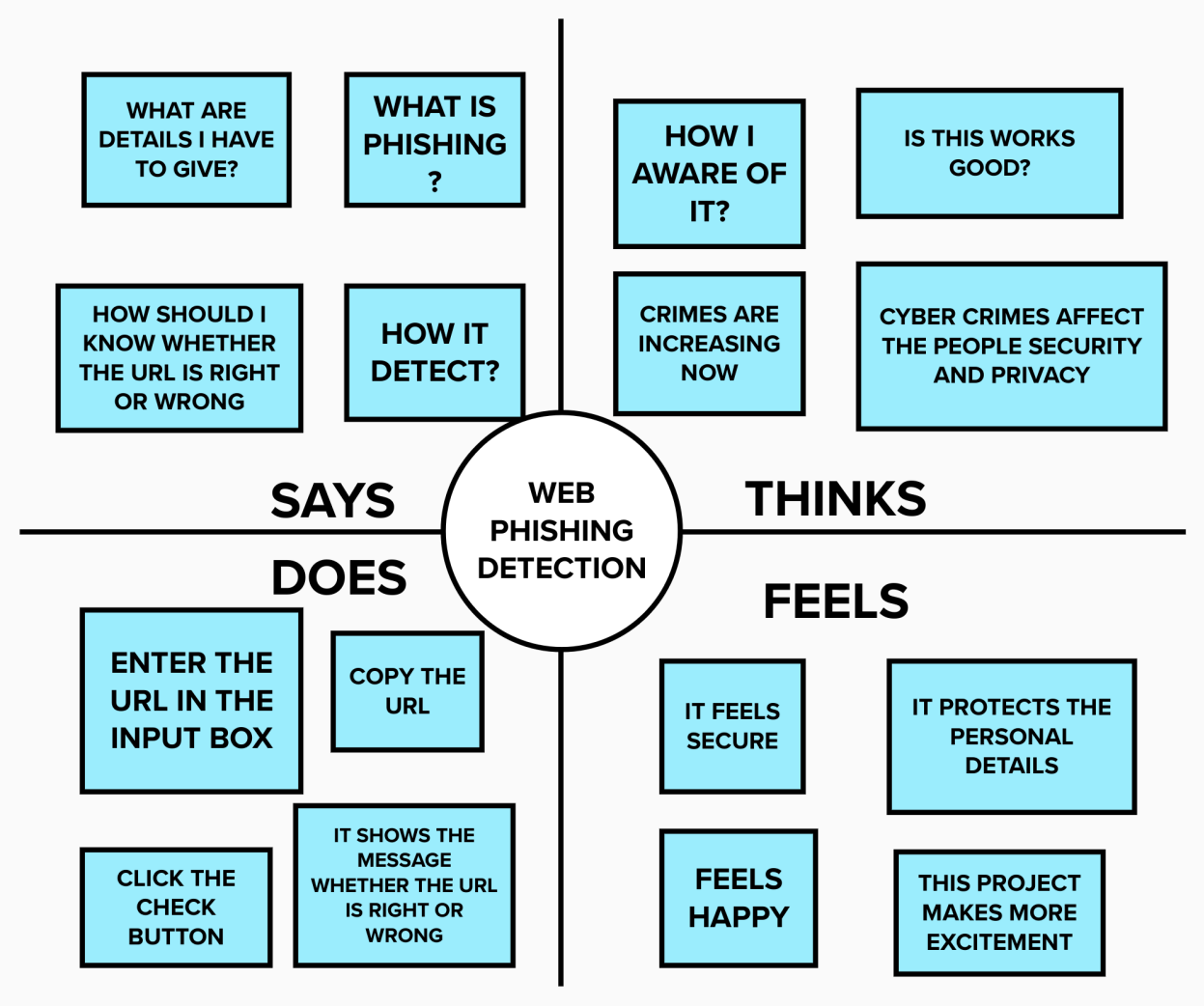
Wang et al., Jain and Gupta and Han et al.use white list-based method for the detection of suspected URL. Blacklist-based methods are widely used in openly available anti-phishing toolbars, such as Google safe browsing, which maintains a blacklist of URLs and provides warnings to users once a URL is considered as phishing. Prakash et al. proposed a technique to predict phishing URLs called Phishnet. In this technique, phishing URLs are identified from the existing

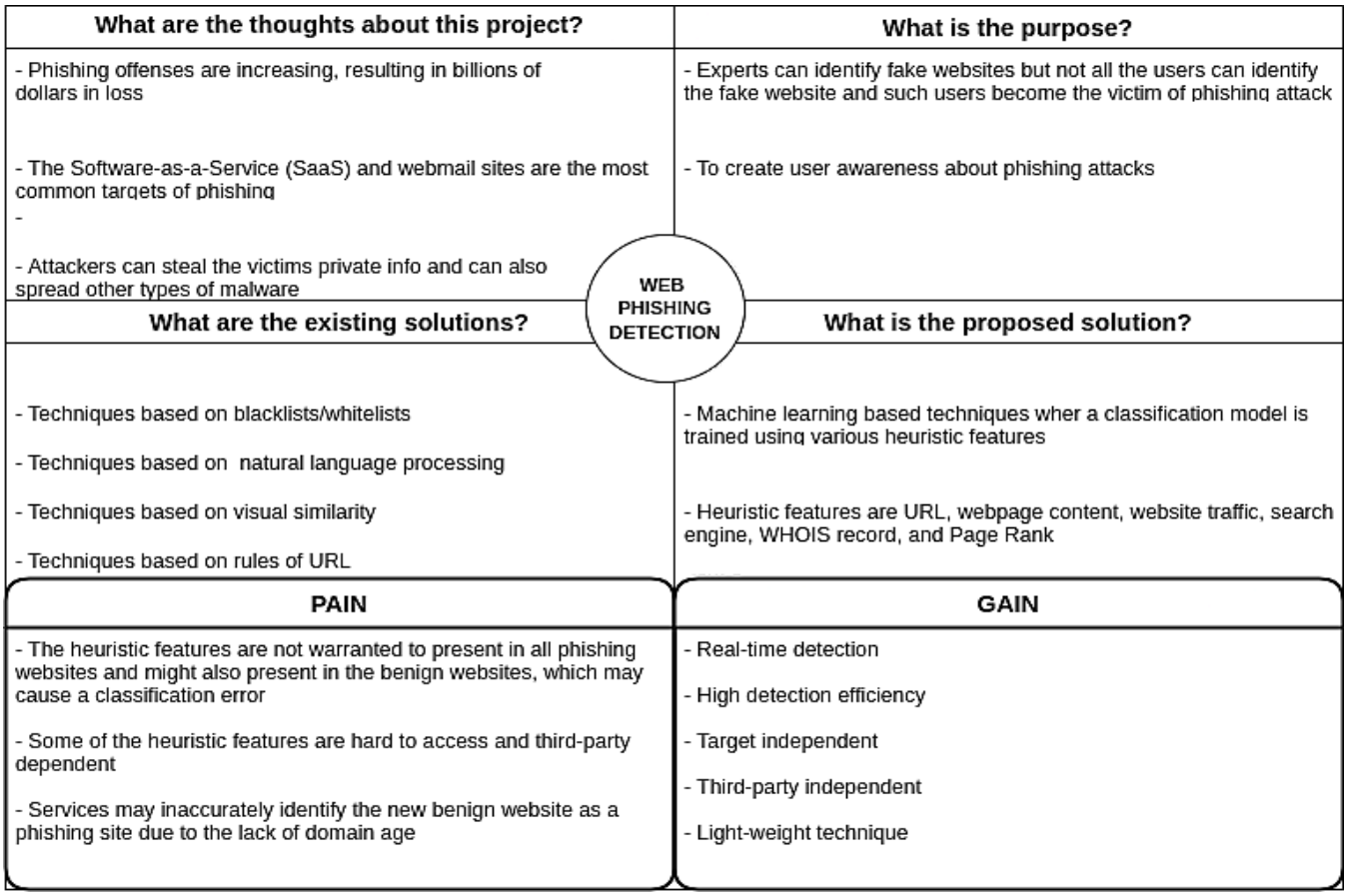
blacklisted URLs using the directory structure, equivalent IP address, and brand name.

Felegyhazi et al. developed a method that compares the domain name and name server information of new suspicious URLs to the information of blacklisted URLs for the classification process. Sheng et al. demonstrated that a forged domain was added to the blacklist after a considerable amount of time, and approximately 50–80% of the forged domains were appended after the attack was carried out.

Prepare Empathy Map

In this activity you are expected to prepare the empathy map canvas to capture the user Pains & Gains, Prepare list of problem statements.





Goal

In order to detect and predict e-banking phishing websites, we proposed an intelligent, flexible and effective system that is based on using classification algorithms. We implemented classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy. The e-banking phishing website can be detected based on some important characteristics like URL and domain identity, and security and encryption criteria in the final phishing detection rate. Once a user makes a transaction online when he makes payment through an e-banking website our system will use a data mining algorithm to detect whether the e-banking website is a phishing website or not.

Ideation

In this activity you are expected to list the ideas (atleast 4 per each team member) by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.

1. Use anti-phishing protection and anti-spam software to protect yourself when malicious messages slip through to your computer.
2. Anti-spyware and firewall settings should be used to prevent phishing attacks and users should Protect your mobile phone by setting software to update automatically. These updates could give you critical protection against security threats.
3. Protect your data by backing it up. Back up your data and make sure those backups aren't connected to your home network. You can copy your computer files to an external hard drive or cloud storage. Back up the data on your phone, too.
4. The website may be used to hack and misuse others detail so to protect that Then kids nowdays are learning from online so to protect them from facing any unpleasant or bad activity. So create a extension in google which will detect the fake websites.
5. Every time you click on a link, look at the browser bar and see if matches exactly the one you would type in to go to your account.
6. All members of your executive and management team are vulnerable. If a phishing scammer acquires the email credentials of high-profile leadership, it's likely they'll target anyone they can using that very email address.
7. Almost all spam messages are malicious emails sent by unknown sources. These sources could be hackers who aim to hack into the computers of their victims.
8. Never respond to spam messages because through this, the spammer will know that the email address is active and thus, it increases the chance of your email to be constantly targeted by the spammer.
9. Do not use your personal or business email address when registering in any online contest or service such as applications, deal updates, etc. Many spammers watch these groups or emailing lists to harvest new email addresses.
10. In fact, many unsuspecting users have been dupped via text message phishing (also known as smishing) and through social media.

11. The threat of malicious messages luring users to click on a link, open a malicious webpage, download malware or provide credentials on a spoofed site proves that threat actors are getting continuously creative in their methods to hijack your assets and steal your credentials.
12. While these attacks use electronic written words to lure a user into their scam and some of the messages may be hosted in social media, a new form of messaging attacks are emerging via other cloud and SaaS (software as a service) platforms that provide in-application messaging between users.

TOP 4:

1. We would create an interactive and responsive website that will be used to detect whether a website is legitimate or phishing. This website is made using different web designing languages which include HTML, CSS, Javascript and Python.
2. It must be noted that the website is created for all users, hence it must be easy to operate with and user-friendly.
3. The website will show information regarding the services provided by us. It also contains information regarding ill- practices occurring in todays technological world.
4. The website will be created with an opinion such that people are not only able to distinguish between legitimate and fraudulent website, but also become aware of the mal-practices occrring in current world. They can stay away from the people trying to exploit ones personal information, like email address, password, debit card numbers, credit card details, CVV, bank account numbers.