**Project Design Phase-I**

**Solution Architecture**

| | |
|---|---|
| Date | 19 September 2022 |
| Team ID | PNT2022TMID15184 |
| Project Name | Project – Web Phishing Detection |
| Maximum Marks | 4 Marks |

## Solution Architecture:

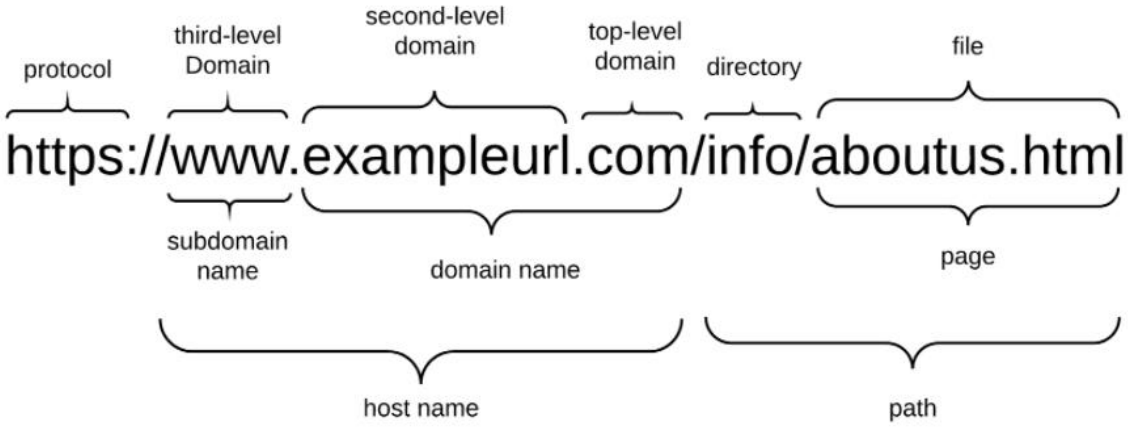
Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

- Find the best tech solution to solve existing business problems.
- Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
- Define features, development phases, and solution requirements.
- Provide specifications according to which the solution is defined, managed, and delivered.

Phishing has become more damaging nowadays because of the rapid growth of internet users. The phishing attack is now a big threat to people's daily life and to the internet environment. In these attacks, the attacker impersonates a trusted entity intending to steal sensitive information or the digital identity of the user, e.g., account credentials, credit card numbers and other user details. A phishing website is a website which is similar in name and appearance to an official website otherwise known as a spoofed website which is created to fool an individual and steal their personal credentials. So, to identify the websites which are fraud, this paper will discuss the machine learning and deep learning algorithms and apply all these algorithms on our dataset and the best algorithm having the best precision and accuracy is selected for the phishing website detection. This work can provide more effective defences for phishing attacks of the future.

The main reason is the lack of awareness of users. But security defenders must take precautions to prevent users from confronting these harmful sites. Preventing these huge costs can start with making people conscious in addition to building strong security mechanisms which are able to detect and prevent phishing domains from reaching the user.

Uniform Resource Locator (URL) is created to address web pages. The figure below shows relevant parts in the structure of a typical URL.




It begins with a protocol used to access the page. The fully qualified domain name identifies the server who hosts the web page. It consists of a registered domain name (second-level domain) and suffix which we refer to as top-level domain (TLD). The domain name portion is constrained since it has to be registered with a domain name Registrar. A Host name consists of a subdomain name and a domain name.An phisher has full control over the subdomain portions and can set any value to it. The URL may also have a path and file components which, too, can be changed by the phisher at will. The subdomain name and path are fully controllable by the phisher. We use the term FreeURL to refer to those parts of the URL in the rest of the article.

The attacker can register any domain name that has not been registered before. This part of URL can be set only once. The phisher can change FreeURL at any time to create a new URL. The reason security defenders struggle to detect phishing domains is because of the unique part of the website domain (the FreeURL). When a domain detected as a fraudulent, it is easy to prevent this domain before an user access to it.

Some threat intelligence companies detect and publish fraudulent web pages or IPs as blacklists, thus preventing these harmful assets by others is getting easier.

Features of phishing

- URL-Based Features
- Domain-Based Features
- Page-Based Features
- Content-Based Features

Development phase: We can utilize a different encryption scheme; asymmetric algorithms can be deployed. We have deployed symmetric encryption due to the efficiency and processing time outperforming asymmetric encryption algorithms. We can even change the symmetric encryption algorithm to something else, like DES, triple DES, or Blowfish. Researchers will be able to test and measure performance for any replaced encryption algorithm. Also, usage of the encryption keys can be change to reflect a different key size for each importance level assigned. For example, we can assign an encryption key of 192 bits instead of 256 bits for the importance level “High” value.

# Example - Solution Architecture Diagram:

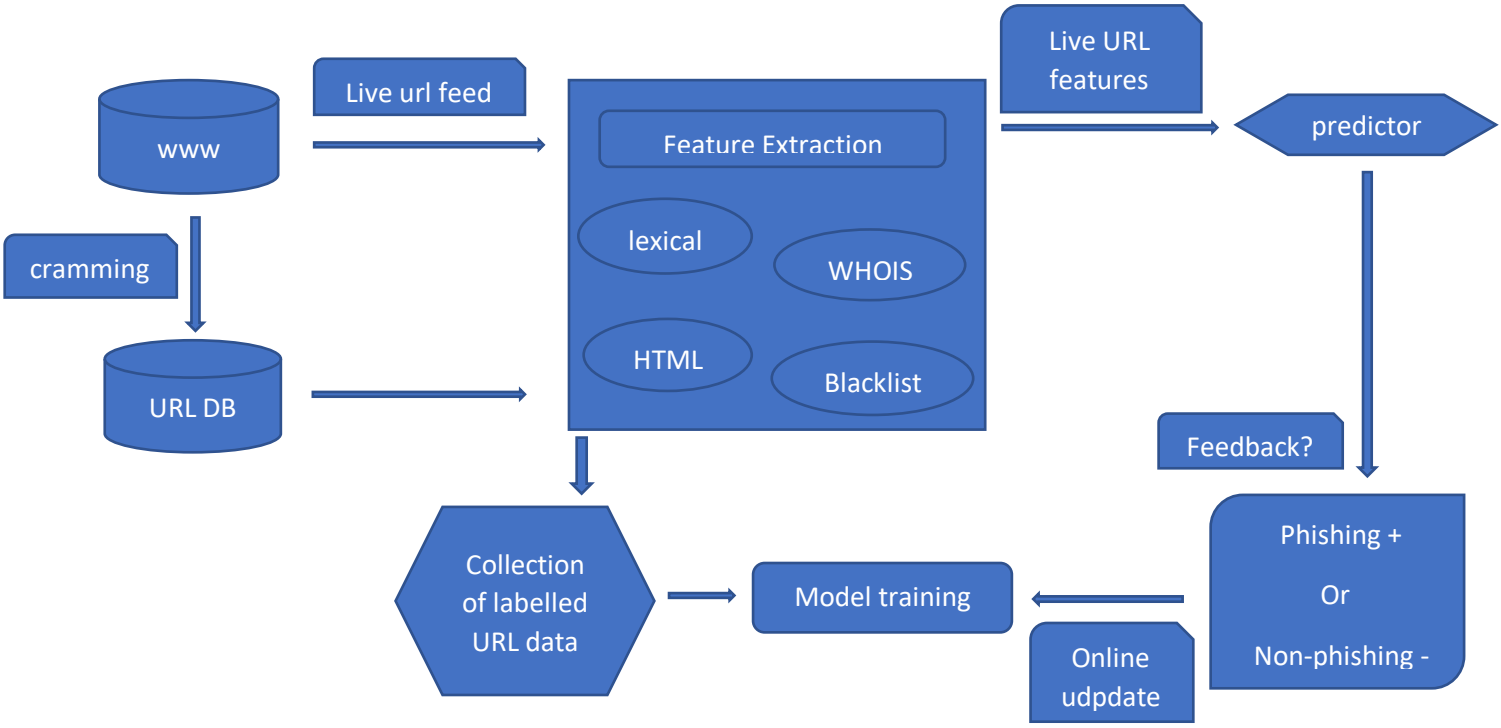



*Figure 1: Architecture and data flow of the web* phishing ***detection***