Project Design Phase-I
Solution Architecture

Date	16 october 2022
Team ID	PNT2022TMID01337
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.

Experts can identify fake websites but not all the users can identify the fake website and such users become the victim of phishing attack. Main aim of the attacker is to steal banks account credentials.The general method to detect phishing websites by updating blacklisted URLs, Internet Protocol (IP) to the antivirus database which is also known as “blacklist" method. To evade blacklists attackers uses creative techniques to fool users by modifying the URL to appear legitimate via obfuscation.

Decision tree algorithm is easy to understand and also easy to implement. Decision tree begins its work by choosing best splitter from the available attributes for classification which is considered as a root of the tree. Algorithm continues to build tree until it finds the leaf node. Decision tree creates training model which is used to predict target value or class in tree representation each internal node of the tree belongs to attribute and each leaf node of the tree belongs to class label. In decision tree algorithm, gini index and information gain methods are used to calculate these nodes.

Some best anti phishing technology in Business:

- Abnormal Security
- Trustifi
- Proofpoint Essentials
- Microsoft Defender for Office 365
- Agari

Describe the structure, characteristics, behavior, and other aspects of the software to project stakeholders.

Phishing is a method used to compromise the computers of and steal sensitive information from individuals by pretending to be an email from or the website of a trusted organization.

For example, a person receives an email that appears to be from the recipient's bank requesting that recipient verify certain information on a web form that mimics

the bank's website. When captured by the hackers, the data allows them access to the recipient's banking information. Alternatively, the web-link may contain malicious code to compromise the target's computer. One of the things that makes phishing attacks tricky is that they can be distributed by compromising the email address books of compromised computers. So the email may appear to have been sent by a known and trusted source.

A subset and highly effective form of phishing attack is a spear-phishing attack in which a hacker will research an intended target and include details in an email that makes the email seem more credible. The details may, for example, reference a corporate social event from the previous month that was published on a public website. It can be exceedingly difficult to protect against these kinds of attacks as demonstrated by the notable and extremely costly breaches of sensitive information by Target, Home Depot, and Baylor Regional Medical Center.

The one thing companies need to keep in mind for phishing attack protection is:

Defending against these attacks requires a coordinated and layered approach to security:

- Train employees to recognize phishing attacks to avoid clicking on malicious links. For example, if the domain of the link to which you are being directed doesn't match the purported company domain, then the link is a fake.
- Many spam filters can be enabled to recognize and prevent emails from suspicious sources from ever reaching the inbox of employees.
- Two factor authentication should be deployed to prevent hackers who have compromised a user's credentials from ever gaining access.
- Browser add-ons and extensions can be enabled on browsers that prevent users from clicking on malicious links.

Features & Methodology:

- Feature Extraction
- Algorithm selection and output classification
- Create baseline model with initial dataset
- Evaluate the performance of the model and fine tuning.
- Apply test data on pre-trained base model and make Prediction

Example - Solution Architecture Diagram:

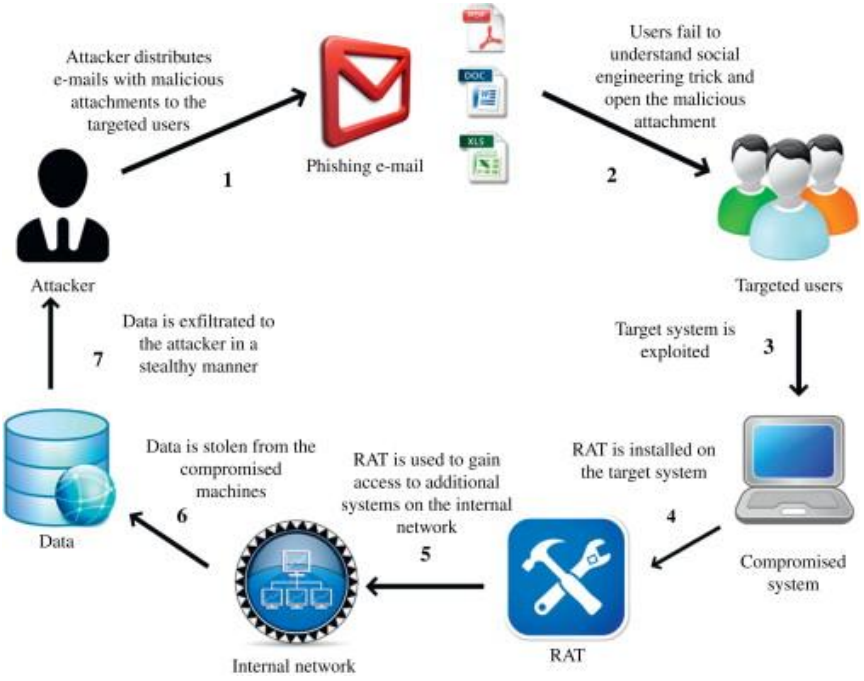

Figure 1: Architecture and data flow to Detect Phishing & Fraud Emails from famous companies.

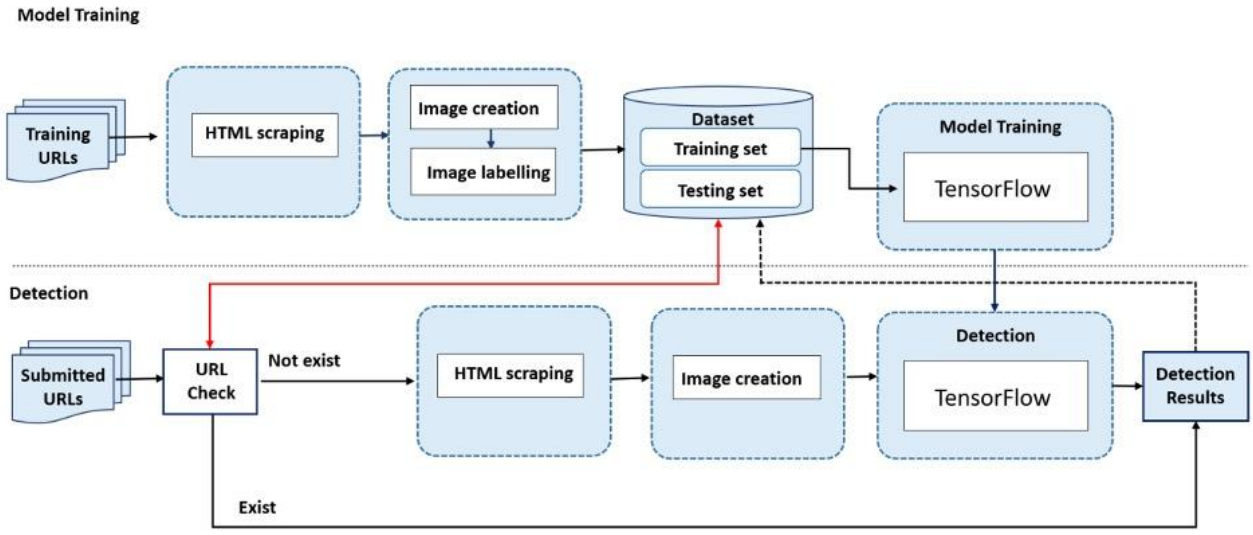

Figure 2: web phishing using Machine Learning.