

BUILDING A MACHINE LEARNING-BASED SOLUTION FOR DETECTING PHISHING LINK

Nhan Le Ton

University of Information Technology
HCMC, Vietnam

What ?

We introduce an efficient and accurate detection system for identifying phishing links, in which we have:

- Proposed a combination of phishing link detection based on static information and visual interface data.
- Developed a web-based system with clearly defined and extensible modules.
- Enhanced the quality of the dataset, ensuring reliability and diversity.

Why ?

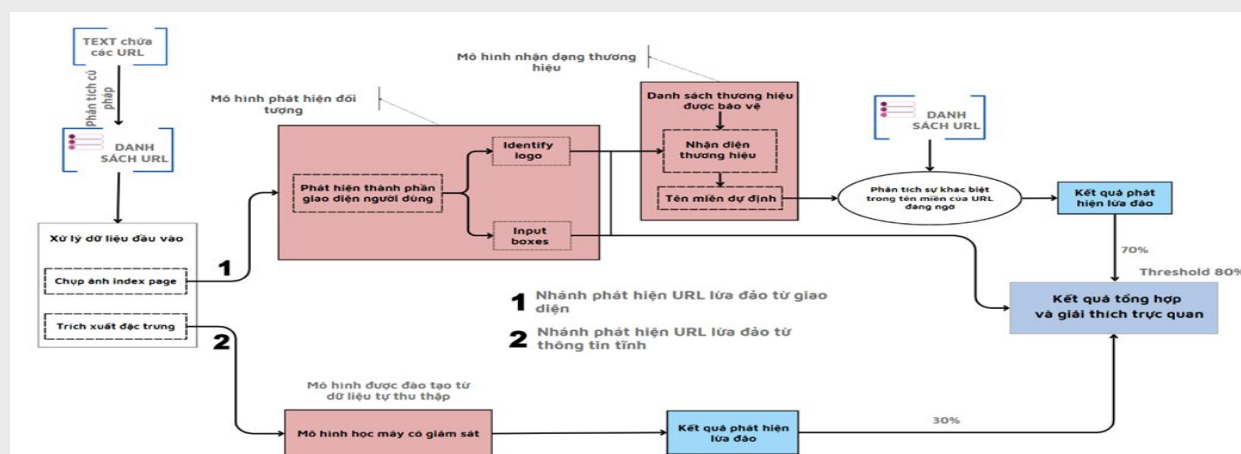
- Phishing attacks cause losses of millions of USD; therefore, early detection of such attacks—especially those involving phishing links—is a critical task.
- Most researchers focus on detecting phishing links using either static information or visual interface data, without combining both approaches.

Overview

Input Data Processing

Phishing Link Detection

Result Aggregation



Description

1. Input Data Processing

- The input data provided by users may consist of text, emails, or SMS messages containing links to be examined. This data will be analyzed to extract a list of URLs.
- Features will then be extracted from the list of URLs, and a screenshot of each index page will be captured.

2. Phishing Link Detection

- Phishing link detection based on interface information involves identifying logos and input boxes on the webpage, followed by brand recognition and generating analytical results.
- Phishing link detection based on static information uses the features extracted during the data processing step, which are then passed through a pre-trained supervised machine learning model to determine whether the link is malicious.

3. Result Aggregation

- The results from the two branches are combined using ensemble techniques such as voting (majority voting or weighted voting) or probability-based methods (average probabilities or weighted average). The final decision is then presented along with detailed results and feature information related to the link on the web interface.

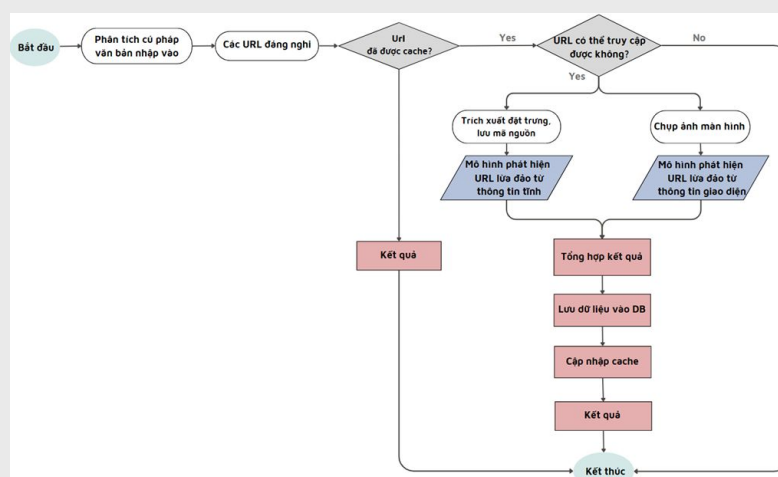


Figure 1: Workflow of the Web Application

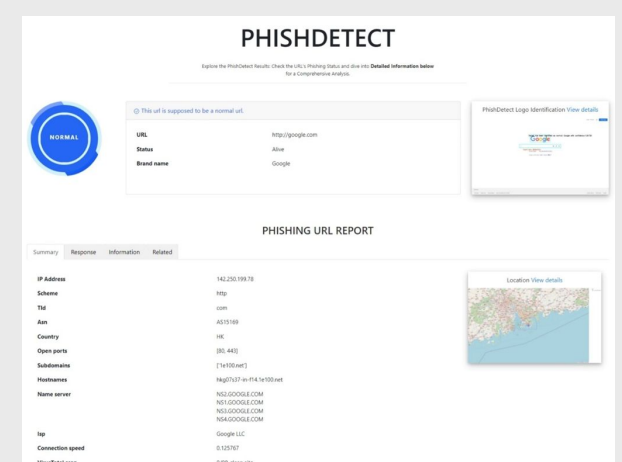


Figure 2: Illustration of the Main Results Page on the Web Interface