Web phishing detection

IDEATION

Team Leader

ABULHASAN A

Team Members:

ASIKAHAMED Y,

AJITHKUMAR B,

GOKURAM D.

Problem statement

Phishing has a list of negative effects on a business, including loss of money, loss of intellectual property, damage to reputation, and disruption of operational activities. These effects work together to cause loss of company value, sometimes with irreparable repercussions.

A common way to obtain phishing detection measurements is to perform an assessment. To continue with the phishing detection example, a measurement of how many suspicious e-mails were reported to security would be collected at the end of each phishing assessment. If you're using a commercial tool, the number of e-mails sent during each assessment is available from the reporting screen.

There are three phases to approaching the phishing attack

Phase 1 focuses on dataset gathering, pre-processing, and feature extraction. The objective is to process data for use in Phase 2. The gathering stage is done manually by using Google crawler and Phish tank, each of these data gathering methods was tested to ensure a valid output. The dataset is validated first after gathering, then normalized, features extracted and finally dataset division.

Phase 2 focuses on designing and implementing training and validating models using a single classifier. Predefined performance metrics are used as a measurement of accuracy, precision, recall, and f-measure. The objective of this phase is to test the performance of individual classifiers.

Phase 3 which corresponds to the third objective is divided into two parts, one is the ensemble design and the other is the comparative study between the best ensemble and the best individual classifier that was selected in Phase 2. To design a good ensemble, only three algorithms are used for individual ensembles due to the selection of majority voting as the ensemble algorithm, an odd number of algorithms must be used to select the committee of ensembles.

This study focuses on investigating a better detection approach and designing an ensemble of classifiers suitable to be used in phishing detection. Summarizes the design and implementation phases leading to the proposed better detection model.