A Detailed Study Phishing Sites via Site Deconstruction & Clustering Techniques

Karthika Subramani, Dr. Roberto Perdisci, Assoc. Professor @ UGA, Dr. Babak Rahbarinia, Asst. Professor @ AUM

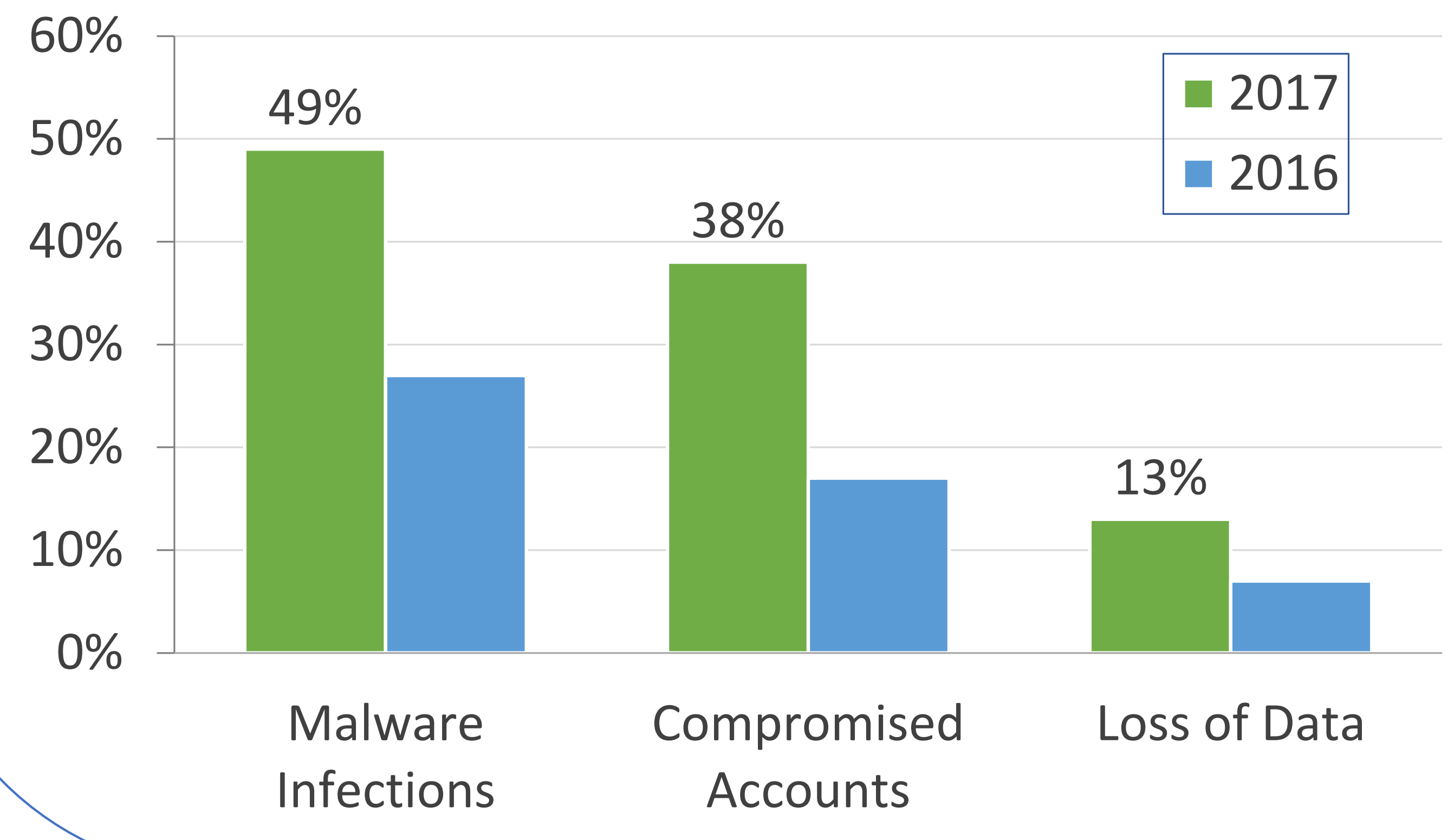
Abstract

Phishing is the top threat vector for cyberattacks. Our system gathers in-depth information of the phishing sites and aims to answer the following questions hoping it might lead to curb the growth of phishing attacks at a greater rate.

- What information is gathered by an attacker?
- What kind of approaches are used in phishing?
- Is there a relation between different phishing sites?
- Which parameters can be used to distinguish a phishing site from a legitimate site?

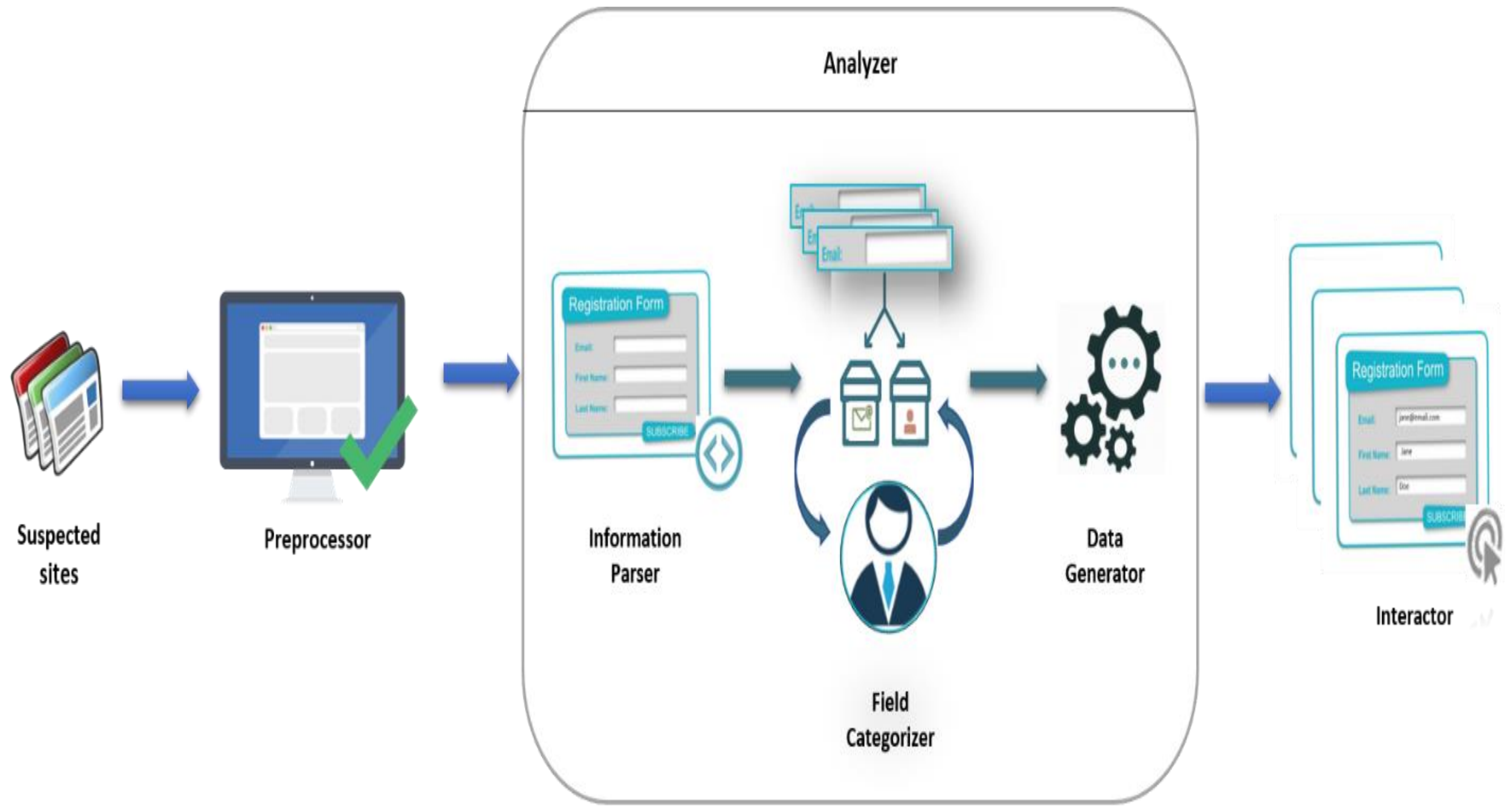
Introduction

Phishing is the attempt to obtain sensitive information by disguising as a trustworthy entity. Phishing attacks are on the rise over past few years.



System Overview

Our system consists of 3 major components that successfully help it to traverse through multiple pages of a phishing site. We obtain suspicious URLs from the PhishTank repository.



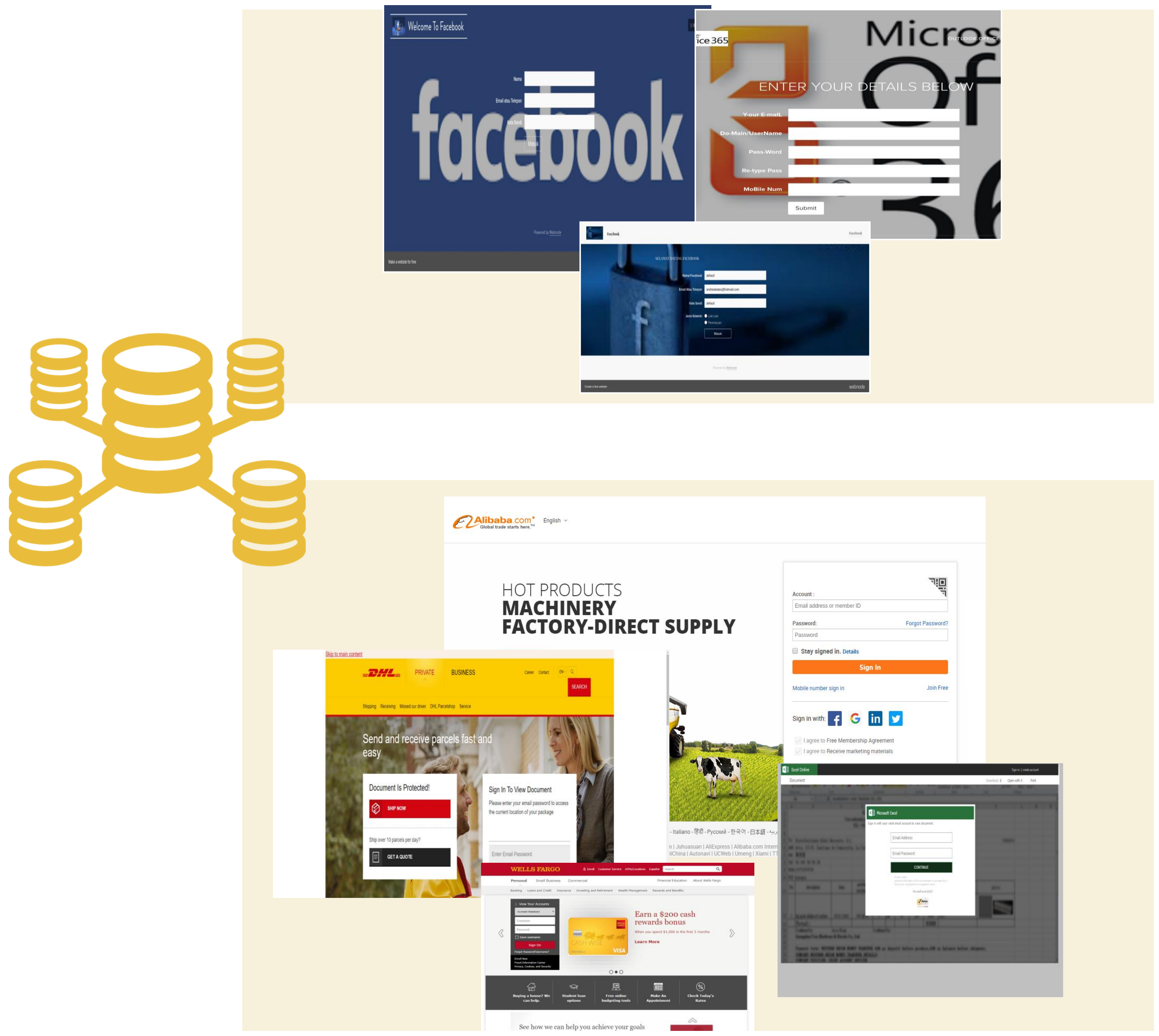
Clustering & Classification

The different techniques used to process the data and the results obtained are as follows

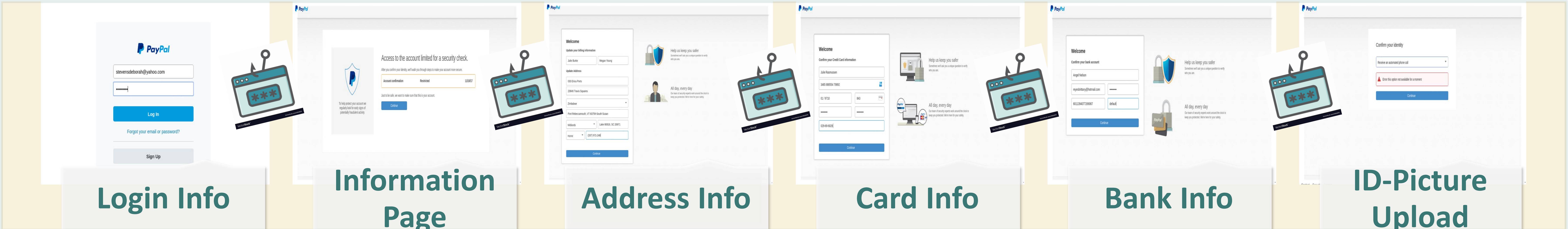
- Classification of input fields with Active Learning
- Hierarchical Clustering for clustering similar phishing sites
- Distinguish benign sites from phishing sites
- Phishing kits created by same developer
- Distribution of the phishing kits

Experiments & Results

- We were able to traverse through multiple pages in most of the sites
- We learnt most sites navigate users to targeted benign sites after stealing credentials from the user
- We were able to cluster sites targeting different domains sharing same properties in their design structure and resource requests



Network Security Intelligence and LAB


PayPal
Login Info
Information Page
Address Info
Card Info
Bank Info
ID-Picture Upload