Informatics Institute of Technology, Sri Lanka

In Collaboration with

University of Westminster, UK

Detection of Social Engineering Attacks: Data Phishing

A dissertation by

Shehan Liyanage

Supervised by

Ms. Sapna Kumarapathirage

Submitted in partial fulfilment of the requirement for

Bachelor of Science (Honours) degree in Computer Science

Department of Computing

2021

Abstract

Social Engineering is considered one of the most renowned and considered one of the easiest ways to implement an attack against computer systems in the cybersecurity domain. This is mainly due to the rapid advancements in digital communication technologies where it has become more accessible and easier to communicate between humans. The foundation of SE attacks is human weaknesses. Due to the availability of personal and sensitive information through online social networking platforms and services, consumers have become much more vulnerable to such malicious social engineering attacks in recent years. With parallel improvements in the Artificial Intelligence domain, considering Machine Learning which is a subset, automated SE attacks have become a new trend among SE attackers.

Data Phishing is a subset of Social Engineering and it is one of the most known methods of attack. The motive of such attacks is to obtain the personal and sensitive information of a user by misleading them psychologically with legitimate-looking e-mails, website links, contact detail forms, shopping site checkouts, customer chat applications, etc. Detecting such phishing techniques is not an easy task as before since the technology used by phishing engineers have evolved considerably. There have been certain studies carried out by researchers to detect phishing attacks with a high rate of accuracy in recent times using Machine Learning techniques such as Natural Language Processing and Artificial Neural Networks. Although the existing systems have been able to generate highly accurate detection results with real as well as semi-synthetic datasets along with different machine learning algorithms.

This research mainly focuses on a system where data phishing attack detection accuracy is increased reasonably as well as zero-day detection, which is the efficiency of the detection of malicious attacks, by using natural language processing and ML algorithms along with datasets. Further, the system is compared with other existing systems to determine whether any major or slight improvement has been accomplished in the process.

Keywords: Social Engineering, Data phishing, Cybersecurity, Machine Learning