SCHOOL OF COMPUTING
Faculty of Engineering

Project Proposal Form MCST1043
Sem: 2 Session: 2024/25

SECTION A: Project Information.

Program Name: **Masters of Science (Data Science)**

Subject Name: **Project 1 (MCST1043)**

Student Name: Cui Zhiwen

Metric Number: MSC241040

Student Email & Phone: cuizhiwen@graduate.utm.my & +60135010819

Project Title: DeepPhish-X: Multi-Modal Feature Engineering for Phishing Detection Using Hybrid Models of Computer Vision, Natural Language Processing, and Graph Neural Networks

Supervisor 1:

Supervisor 2 / Industry Advisor(if any):

SECTION B: Project Proposal

Introduction:

The cybersecurity threats facing enterprises are becoming increasingly complex.

Attacker:Build and maintain attack infrastructure for a long time,Design Trojan virus, and continuously improve and fight,Participate in many attack incidents,Plan carefully before attack, conduct multiple tests .

As the means of cyber attacks gradually increase, data analysis can help companies that have not been attacked to take adequate measures to protect themselves.

Problem Background:

With the rapid development of big data, the Internet of Things, and cloud computing, increasingly fierce cyber attacks have become a new challenge to enterprise security. Enterprises may be threatened by cyber attacks at any time.

In this context, we use data analysis to find out the most vulnerable ways for enterprises to be attacked.

Traditional vulnerability detection technologies (such as regular expression matching) have a high false positive rate, while machine learning (such as hidden Markov models) and syntax tree analysis (such as AST-Probe) have become emerging solutions.

Problem Statement:

Discover common methods of corporate network attacks through data analysis

Enterprise security vulnerability management is a core issue in network security. According to the analysis of Webpage 4, vulnerability exploitation has become the "dark battlefield" of cyber attacks between countries. As the main target of attacks, enterprises are in urgent need of data-driven vulnerability analysis tools12.

Data science has significant advantages in vulnerability detection (such as static code analysis, dynamic behavior modeling) and defense strategy optimization (such as vulnerability prioritization). For example, the VDet for Java and AUSERA tools mentioned above demonstrate the high accuracy of deep learning in automated vulnerability detection

Aim of the Project:

The project aims to discover common methods of corporate network attacks through data analysis.

In order to :

Helping businesses that have not been attacked by cyber attacks to protect themselves

Reduce the number of companies that are successfully attacked by cyber attacks

3. Combine internal enterprise logs with external threat intelligence data to build a hybrid data model to improve detection accuracy

Objectives of the Project:

Analyze common network attack methods and network defense methods

Identify which businesses are most vulnerable

Explore blockchain technology for attack tracing or federated learning analysis methods to protect data privacy

Scopes of the Project:

Data from some companies in Malaysia Or China that were attacked by cyber attacks within one year April 2024 to April 2025

Log data: firewall logs, intrusion detection system (IDS) logs, terminal device logs, etc.;Network traffic data: Capture protocol packets (such as DNS, HTTP) through traffic mirroring for behavioral analysis;Threat intelligence data: Integrate public threat databases (such as MITRE ATT&CK) to assist in attack pattern identification.

Expected Contribution of the Project:

Visualize the network attacks that enterprises are vulnerable to, and optimize defense measures through modeling.

Design a real-time attack detection framework based on streaming data processing (such as Apache Kafka)

Project Requirements:

Software: Phyton, Power BI and Excel

Hardware:

Technology/Technique/ Methodology/Algorithm: Exploratory Data Analysis, Visualisation

Type of Project (Focusing on Data Science):

[ ] Data Preparation and Modeling

[ ] Data Analysis and Visualization

[ ] Business Intelligence and Analytics

[ ] Machine Learning and Prediction

[ ] Data Science Application in Business Domain

Status of Project:

[ ] New

[ ] Continued

If continued, what is the previous title? .....

SECTION C: Declaration

I declare that this project is proposed by:

[ ] Myself .....

[ ] Supervisor/Industry Advisor ( ) .....

Student Name: .....

....

Signature **Date**

SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

I/We agree to become the supervisor(s) for this student under aforesaid proposed title.

Name of Supervisor 1: .....

....

Signature **Date**

Name of Supervisor 2 (if any): .....

....

Signature **Date**

SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

Result:

[ ] FULL APPROVAL [ ] CONDITIONAL APPROVAL (Major)*

[ ] CONDITIONAL APPROVAL (Minor) [ ] FAIL*

***** Student has to submit new proposal form considering the evaluators' comments.

Comments:

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Name of Evaluator 1: .....

Signature **Date**

Name of Evaluator 2: .....

Signature **Date**