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

Program Name:	**Masters of Science (Data Science)**
Subject Name:	**Project 1 (MCST1043)**
Student Name:	Cui ZhiWen
Metric Number:	MCS241040
Student Email & Phone:	cuizhiwen@graduate.utm.my
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):	

Abstract

This report introduces **DeepPhish-X**, a novel system designed for phishing webpage detection, a critical task in cybersecurity with significant implications for online safety and data protection. Traditional methods, primarily relying on URL feature analysis, are limited in capturing the full context of sophisticated phishing attacks.[1] DeepPhish-X addresses this by proposing an innovative multi-modal approach that combines HTML Document Object Model (DOM) graph modeling with URL feature analysis, leveraging advanced deep learning techniques.

DeepPhish-X utilizes Graph Convolutional Networks (GCNs), a core component of **Graph Neural Networks**, to model the intricate structure of HTML DOM graphs. Simultaneously, it integrates Convolutional Neural Networks (CNNs), drawing on principles from **Computer Vision** for character-level URL analysis, and Transformer Networks, a powerful tool in **Natural Language Processing**, for capturing word sequence features of URLs.[1] These diverse multi-modal features are then integrated using another Transformer network, which excels at selectively capturing the interdependencies and complementary relationships between these distinct feature sets.[1]

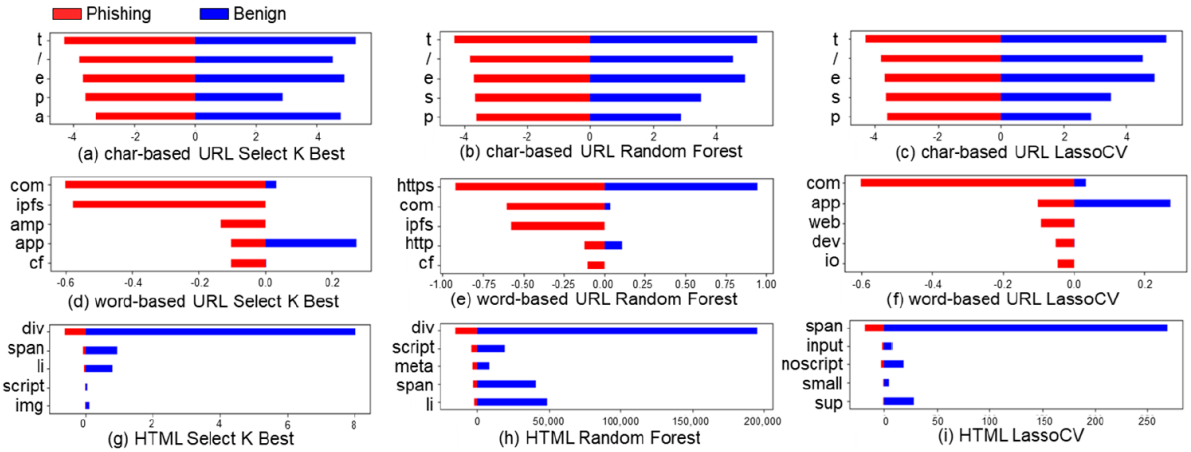
DeepPhish-X was rigorously evaluated on a real-world dataset comprising URL and HTML DOM graph data collected from 2012 to 2024, which includes over 80 million nodes and edges, providing a robust foundation for testing.[1] Experimental results demonstrate a significant improvement in classification accuracy by 7.03 percentage points compared to existing state-of-the-art techniques.[1] Additionally, ablation experiments further validated the effectiveness of individual features within the DeepPhish-X model.[1] These results confirm the efficacy of integrating HTML DOM structure and URL features using a hybrid deep learning approach. DeepPhish-X significantly enhances phishing detection capabilities, offering a more accurate and comprehensive solution for identifying malicious webpages.[1]

1. Introduction

The rapid expansion of internet services has revolutionized how individuals and organizations communicate, conduct transactions, and access information.[1] However, this growth has also led to an increase in cybersecurity threats, with phishing attacks becoming one of the most

widespread and serious forms of online fraud. Phishing attacks involve creating fake websites that appear to be benign, in order to steal sensitive information such as passwords, credit card numbers, and personal identification details from users.[1] As phishing techniques become increasingly sophisticated, detecting these fraudulent activities has become more challenging.[1]

Traditional phishing detection methods primarily rely on URL feature analysis, but this approach is often insufficient to capture the full context of phishing attacks. For example, phishing websites may use domains or paths similar to legitimate websites to deceive users, making detection difficult solely through URL analysis.[1] In such cases, attackers prioritize superficial imitation (like URLs) to trick human users and simple detection systems. However, despite URLs being imitable, phishing websites rarely fully replicate the HTML structure of legitimate websites.[1] This inherent complexity of legitimate website HTML structures makes them difficult to perfectly copy or obfuscate, creating a fundamental contradiction in phishing attack design. This inherent difficulty in replicating full structural integrity, rather than just URL imitation, becomes a core vulnerability that DeepPhish-X aims to exploit. This means effective phishing defense must go beyond superficial indicators, revealing the tell-tale signs of fraudulent websites through machine-driven deep analysis of the underlying webpage structure. This suggests that future cybersecurity research should focus more on features that are computationally intensive or difficult for attackers to perfectly forge, thereby raising the bar for successful phishing campaigns.



This study introduces **DeepPhish-X**, a novel multi-modal approach that significantly enhances phishing detection accuracy by combining URL analysis and HTML DOM structure analysis.[1] DeepPhish-X leverages Graph Convolutional Networks (GCNs), a key component of

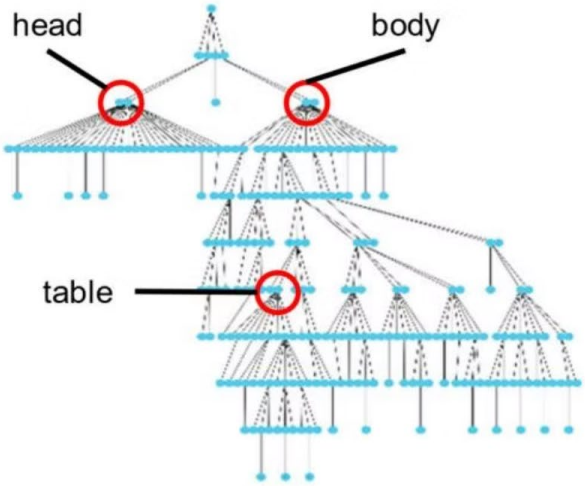
Graph Neural Networks, to model the complex dependencies within the DOM structure. It also employs Convolutional Neural Networks (CNNs), inspired by **Computer Vision** techniques for

processing grid-like data, and Transformer Networks, a powerful architecture from **Natural Language Processing**, to analyze URL features at both the character and word levels.[1] This multi-modal integration allows DeepPhish-X to capture both the structural and sequential characteristics of phishing websites, making it more robust against sophisticated phishing techniques.[1]

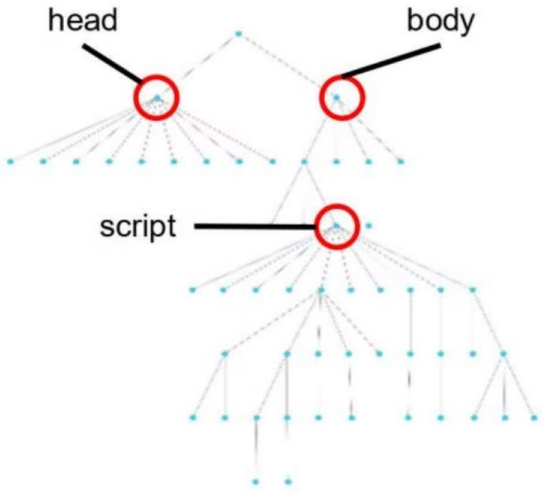
To illustrate the effectiveness of DeepPhish-X, Figure 1 visually demonstrates how key features, extracted using different models, distinguish between benign and phishing websites.[1] Specifically, Figure 1a-c shows URL character-level analysis (akin to image processing in Computer Vision), Figure 1d-f shows URL word-level analysis (a Natural Language Processing task), and Figure 1g-i depicts HTML tag name analysis (contributing to structural understanding for Graph Neural Networks). These features are critical inputs to the DeepPhish-X deep learning model and play a pivotal role in accurately detecting phishing websites.[1] This visualization underscores the importance of these selected features in distinguishing between benign and phishing websites, highlighting the model's superior detection capability. This data-driven argumentation, rather than purely theoretical derivation, strengthens the rationale for the complexity of the proposed model architecture. It emphasizes an important methodological principle in technical research: complex solutions are best justified by clear empirical observations of the problem's nuances.

Furthermore, DeepPhish-X demonstrates that phishing websites, while often mimicking the URL patterns of legitimate websites, fail to effectively replicate the HTML structure.[1] For instance, as shown in Table 1 and Figure 2, both URLs follow a similar pattern of "script.google.com", but the HTML DOM structures differ significantly. Case (a) represents a well-structured benign website, whereas Case (b) is a phishing website with a more simplified and irregular structure. This difference in HTML complexity provides crucial information for DeepPhish-X to identify phishing websites, even when the URL appears legitimate.[1]

DeepPhish-X has been validated using a large-scale real-world dataset, achieving a 7.03% improvement in classification accuracy compared to existing state-of-the-art techniques.[1] These results demonstrate that a deep learning-based approach, which combines HTML DOM structure and URL features, plays a crucial role in enhancing phishing detection capabilities.[1]



(a)



(b)