Phishing Website Detection Method Based on Multimodal Large Language Models

Abstract

Phishing websites continue to pose a serious cybersecurity threat by imitating legitimate online services to deceive users into revealing sensitive information. Traditional phishing detection approaches, such as blacklist-based filtering and machine learning models relying on handcrafted features, are increasingly ineffective against modern phishing attacks that employ obfuscation, dynamic content, and short-lived domains. To address these limitations, this abstract paper presents a phishing website detection approach based on multimodal large language models (LLMs).

The proposed approach utilizes multiple modalities of website information, including URL characteristics, HTML source code features, visual webpage content, and security-related metadata such as SSL certificate and domain information. These heterogeneous features are transformed into structured textual representations, enabling a large language model to perform multimodal feature fusion through semantic understanding. By reformulating phishing website detection as a text classification problem, the model leverages the reasoning capability of LLMs to identify phishing patterns more effectively.

The expected outcome of this approach is improved phishing detection accuracy and better generalization to previously unseen phishing websites when compared to traditional single-modality methods. This abstract paper serves as a preliminary proposal for a full IEEE-style base paper and subsequent project implementation.

Keywords: Phishing Website Detection, Multimodal Learning, Large Language Models, Cybersecurity