Notification No.: 579/2025

Date of award: 01-05-2025

Name of the Scholar: Akash Shah

Student ID: 202009029

Name of the Supervisor: Prof. (Dr.) Monica Mehrotra

Name of the Department: Department of Computer Science

Thesis Title: Malicious User Identification and their impact on Legitimate Users in Online Social Network Platforms

Key Findings of the Thesis

This thesis conducts an in-depth investigation into the identification of malicious users on Online Social Network (OSN) platforms and introduces four novel deep learning-based frameworks, each designed to address a distinct category of threat. Through extensive experimentation and evaluation on real-world datasets, the study demonstrates the efficacy of the proposed models in enhancing detection accuracy, supporting effective classification, and reinforcing privacy preservation mechanisms.

This thesis investigates the presence and impact of malicious users on Online Social Network (OSN) platforms and proposes four deep learning-based models, each addressing a specific challenge. Through extensive experimentation and validation on real-world datasets, the study demonstrates the efficacy of the proposed models in enhancing detection accuracy, supporting effective classification, and reinforcing privacy preservation mechanisms. The key findings outlined below highlight the overall impact and effectiveness of each proposed method.

GRUbBD effectively detects bots by learning behavioral patterns and profile metadata using GRU with attention. Its combination of TF-IDF embeddings and LDA-based dimensionality reduction ensures both accuracy and computational efficiency across platforms like Twitter and Instagram.

DeepMUI demonstrates that handling heterogeneous user data (textual, numerical, binary) enhances malicious user identification. Its hybrid CNN-LSTM architecture, Word2Vec embeddings, and a custom pooling layer result in robust detection of fake and automated profiles.

DeepURLGuard outperforms traditional models in classifying URLs (benign, phishing, malware, defacement) using lexical, content, and network features. Bi-RNN with attention, PSO optimization, and XAI-based interpretability together offer both accuracy and transparency.

DepInferAttack confirms that models trained on social data are vulnerable to membership inference attacks. GPT-2-generated synthetic data is effective for simulating attacks, and the integration of homomorphic encryption significantly reduces attack success, enhancing data privacy.

Overall, the thesis establishes that specialized deep learning models with attention mechanisms, hybrid designs, and privacy-aware techniques can effectively strengthen OSN security while maintaining scalability and trust.

Thesis Outline

This thesis is structured into seven chapters, each addressing a key aspect of malicious user identification and its implications in Online Social Network (OSN) platforms. It systematically introduces novel deep learning models designed to detect bots, fake profiles, malicious URLs, and defend against privacy breaches like membership inference attacks.

Chapter 1: Introduction

This chapter sets the foundation by discussing the growing influence of Online Social Networks and the associated risks due to malicious entities. It outlines the motivation, problem statement, research objectives, scope, and contributions of the thesis. It also presents the research challenges and provides an overview of the methodologies adopted.

Chapter 2: Literature Review

A comprehensive review of existing studies on malicious users, bot detection, fake profiles, malicious URLs, and inference attacks is presented. The chapter identifies research gaps in behavior-based detection, interpretability, generalizability across platforms, and privacy protection. These gaps directly motivate the development of the models proposed in the subsequent chapters.

Chapter 3: GRUbBD – GRU-based Bot Behavior Detector

This chapter proposes the GRUbBD model, which integrates Gated Recurrent Units (GRUs) with an attention mechanism to detect bots on OSNs using behavioral and textual features. It utilizes TF-IDF for text processing and applies Linear Discriminant Analysis for dimensionality reduction. The model is evaluated on Twitter and Instagram datasets, showing high bot-detection accuracy and reduced computational overhead.

Chapter 4: DeepMUI – Deep Malicious User Identifier

The chapter introduces DeepMUI, a hybrid deep learning model combining CNN and LSTM to detect fake and automated user profiles. It processes mixed-format data (numerical, binary, and textual) and uses Word2Vec embeddings along with a custom pooling layer. Experimental results confirm the model's superior performance in detecting diverse forms of malicious users.

Chapter 5: DeepURLGuard – A Multiclass Malicious URL Classifier

This chapter presents DeepURLGuard, a Bi-RNN-based model enhanced with a modified Bahdanau attention mechanism to classify URLs into four categories: benign, phishing, malware, and defacement. It extracts lexical, content, and network features and incorporates PSO for optimization. The use of Explainable AI further enhances model transparency and reliability in real-world OSN scenarios.

Chapter 6: DepInferAttack – Membership Inference Attack Framework

This chapter explores privacy vulnerabilities by introducing the DepInferAttack model. Using GPT-2 to generate synthetic data, it evaluates the success of membership inference attacks on ML models trained with OSN data. The proposed use of homomorphic encryption demonstrates a significant reduction in attack success, enhancing privacy protection for users.

Chapter 7: Conclusion and Future Work

The final chapter summarizes the major contributions and findings of the research. It reflects on how each proposed model advances the state of the art in malicious user detection and privacy preservation. The chapter concludes by suggesting future directions, including multimodal analysis, real-time implementation, and deeper exploration of adversarial defenses.