NEW YORK INSTITUTE OF TECHNOLOGY

Department of Computer and Engineering Sciences

DTSC-870

Saptharish Golagani
Sri sakthicharan NirmalKumar
Shruthi Manchappanahalli Basavaraj
Siva Satya Dheeraj Varma

Supervisor: Jerry Cheng

A report submitted in partial fulfilment of the requirements of
the University of Reading for the degree of
Master of Science in *Data Science and Advanced Computing*

May 19, 2025

Abstract

This project report presents a comprehensive AI-powered phishing detection system that integrates multimodal analysis using email, URL, and image data. The system leverages traditional machine learning models and BERT for text-based analysis and CNNs for image analysis. The fusion model employs F1-score weighted soft voting to combine model predictions, enhancing detection accuracy. Additionally, a federated learning framework is implemented to facilitate decentralized model training, preserving data privacy and security. This report details the design, implementation, evaluation, and future prospects of the phishing detection system.

Acknowledgements

Acknowledging the support and guidance of Prof. Jerry Cheng and my project teammates Saptharish Golagani, Sri Sakticharan Nirmal Kumar,Shruthi Manchappanahalli Basavaraj and Siva Satya Dheeraj Varma Sreevatsavaya.

Contents

Chapter 1

Introduction

1.1 Background

Phishing is a prevalent cyber-attack method involving fraudulent attempts to obtain sensitive information such as usernames, passwords, and credit card details by disguising as a trust worthy entity in electronic communications. With the increasing sophistication of phishing techniques, traditional single-modal detection systems (e.g., email-only, URL-only) are becoming less effective. This project addresses the limitations of single-modal approaches by developing a comprehensive phishing detection system that integrates multiple data modalities – email content, URLs, and images.

1.2 Problem statement

Existing phishing detection systems predominantly rely on single data modalities, limiting their detection capabilities. Additionally, centralized training of detection models raises concerns regarding data privacy and security. This project aims to develop a multimodal phishing detection system that integrates multiple data types while preserving data privacy through federated learning.

1.3 Aims and objectives

- Develop a multimodal phishing detection system incorporating email, URL, and image analysis.
- Implement a fusion model using F1-score weighted soft voting to enhance prediction accuracy.

- Integrate federated learning to facilitate decentralized model training without exposing raw data.
- Evaluate the effectiveness of the fusion model and federated learning framework using comprehensive performance metrics.

1.4 Solution approach

The proposed system employs a multi-stage approach consisting of three key components:

- Data Collection and Preprocessing: Emails, URLs, and images are collected and preprocessed to extract relevant features.
- Model Development: Individual models are developed for each modality using BERT, CNNs, and traditional ML classifiers. The fusion model aggregates predictions using F1-score weighted soft voting.
- Federated Learning Implementation: A client-server architecture is established to enable federated training of the fusion model, enhancing data privacy and security.

1.5 Summary of contributions and achievements

- Implementation of a comprehensive multimodal phishing detection system integrating email, URL, and image analysis.
- Development of a fusion model utilizing F1-score weighted soft voting for enhanced prediction accuracy.
- Successful implementation of a federated learning framework for decentralized model training.

Chapter 2

Literature Review

Overview of Phishing Detection Techniques: Phishing detection systems have evolved significantly, transitioning from basic content filtering methods to advanced machine learning-based detection systems. Early detection methods relied on heuristic analysis and rule-based systems. Modern approaches include:

- Text-Based Analysis using Natural Language Processing (NLP) and transformers (e.g., BERT, GPT).
- URL Analysis using lexical, content-based, and behavioral features.
- Image Analysis utilizing CNNs and transfer learning to detect embedded phishing logos or misleading visual content.

2.1 Role of Federated Learning in Cybersecurity:

Federated Learning (FL) enables the development of machine learning models across decentralized data sources without exposing raw data. In cybersecurity, FL is instrumental in:

- Preserving data privacy by training models locally and aggregating encrypted model updates.
- Enhancing data security in phishing detection by preventing data leakage.
- Facilitating collaboration across multiple organizations to build robust, generalized phishing detection systems.

2.1.1 Reference Resources

You can find additional referencing resources below:

- https://www.mdpi.com/1424-8220/23/9/4346
- https://arxiv.org/abs/2110.06025v1

2.1.2 Applications of Fusion Learning in Multimodal Systems:

- Fusion learning combines data from multiple sources to improve predictive accuracy. In phishing detection, this approach leverages:
- Email content analysis using BERT and traditional ML classifiers.
- URL feature extraction using SVM, XGBoost, and BERT.
- Image analysis using CNNs (ResNet, EfficientNet, etc.).
- The proposed system employs F1-score weighted soft voting to combine predictions, enhancing overall system accuracy.

2.2 Homomorphic Encryption for Secure Data Processing:

Encryption (HE) allows computations on encrypted data without decryption. In phishing detection, HE ensures:

- Secure aggregation of model predictions in the federated learning framework.
- Protection of sensitive user data during training and prediction phases.
- Prevention of data leakage through encryption of feature vectors and model parameters.

2.3 Summary and Critical Analysis:

The literature demonstrates the potential of combining federated learning, fusion learning, and homomorphic encryption to build secure, accurate, and privacy-preserving phishing detection systems. The proposed system addresses existing gaps by integrating multiple data modalities and implementing FL and HE for robust, secure predictions.

Chapter 3

Methodology

3.1 Data Collection and Preprocessing

Data for the phishing detection system was sourced from multiple datasets comprising email texts, URLs, and phishing images. The data was preprocessed as follows:

- Email Data: Text preprocessing included tokenization, stopword removal, and conversion to lowercase. The data was then vectorized using TF-IDF and BERT embeddings.
- URL Data: URLs were analyzed using lexical features (length, domain age) and content-based features (presence of suspicious keywords). Additionally, BERT was used to extract semantic features.
- Image Data: Images were resized to 224x224 pixels and normalized. Pretrained CNN models (ResNet, EfficientNet, DenseNet) were employed for feature extraction.

3.2 Model Architectures:

- Email Model: BERT for email content classification combined with SVM and Random Forest for traditional ML-based analysis.
- URL Model: BERT for semantic analysis combined with XGBoost, LightGBM, and CatBoost for ensemble learning.
- Image Model: CNN-based classifiers using ResNet, EfficientNet, and DenseNet for phishing image detection.

3.3 Fusion Model - F1-Score Weighted Soft Voting:

The fusion model aggregates predictions from the three modalities using F1-score weighted soft voting. The formula is: Fusion Score = Fusion Score =

$$\frac{F1_{\text{email}} \times P_{\text{email}} + F1_{\text{url}} \times P_{\text{url}} + F1_{\text{image}} \times P_{\text{image}}}{F1_{\text{email}} + F1_{\text{url}} + F1_{\text{image}}}$$

3.4 Federated Learning Setup - Client-Server Communication

The federated learning framework is implemented using Flower (FLWR). Each client hosts local data (emails, URLs, images) and trains individual models. The server aggregates encrypted model updates using weighted averaging and broadcasts the updated model back to clients. Communication is secured using encryption to prevent data leakage.

3.5 Implementation of Homomorphic Encryption:

Homomorphic encryption is employed to secure model predictions during aggregation. Encrypted feature vectors are sent to the server using CKKS scheme from TenSEAL. Decryption occurs only after the aggregated predictions are received by the client

3.6 Model Training and Evaluation

Each model was trained using stratified k-fold cross-validation. Evaluation metrics include Accuracy, Precision, Recall, and F1 Score. The performance of the fusion model is compared with individual modality models to validate the effectiveness of the proposed system.

3.7 System Architecture

Data Ingestion Layer: Collects email, URL, and image data from clients. Preprocessing Layer: Standardizes and extracts features from each modality. Model Layer: Hosts BERT, traditional ML classifiers, and CNN models. Fusion Layer: Aggregates predictions using F1-score weighted voting. Federated Learning Layer: Facilitates model training without central data

storage. Encryption Layer: Secures data and model updates using homomorphic encryption.

Chapter 4

Results

4.1 Email Phishing Detection Model

The email phishing detection model achieved notable performance through the combined use of BERT and traditional ML models (SVM, Random Forest). Performance Metrics: Accuracy: 92.4BERT significantly enhanced detection by capturing contextual nuances in email content, while traditional models provided complementary insights through TF-IDF-based analysis.

4.2 URL Phishing Detection Model

The URL detection model integrated BERT for semantic analysis and XGBoost, LightGBM, and CatBoost for feature extraction and classification. Performance Metrics: Accuracy: 90.1BERT effectively identified malicious URLs based on contextual patterns, while ensemble learning improved overall accuracy.

4.3 Image Phishing Detection Model

The image detection model employed ResNet, EfficientNet, and DenseNet for visual analysis. Performance Metrics: Accuracy: 87.6The CNN models successfully detected phishing images based on logo recognition and suspicious visual patterns

4.4 Fusion Model Implementation in Multi Modal

The fusion model demonstrated effective multimodal phishing detection by integrating predictions from email, URL, and image models. The multimodal approach enabled the system to leverage complementary data features, resulting in a higher detection accuracy (F1 score: 0.93) than individual models.

The fusion model combined the predictions of email, URL, and image models using F1-score weighted soft voting.

- Performance Metrics:
- Accuracy: 93.2
- Precision: 0.91
- Recall: 0.95
- F1 Score: 0.93

The fusion model demonstrated a significant improvement over individual modalities by leveraging complementary data features and F1-score-based weighting. Additionally, the model operates under a conservative approach where if any single modality detects a phishing threat, the entire prediction is flagged as phishing. This approach ensures comprehensive threat detection, reducing the risk of false negatives and reinforcing the overall robustness of the system.

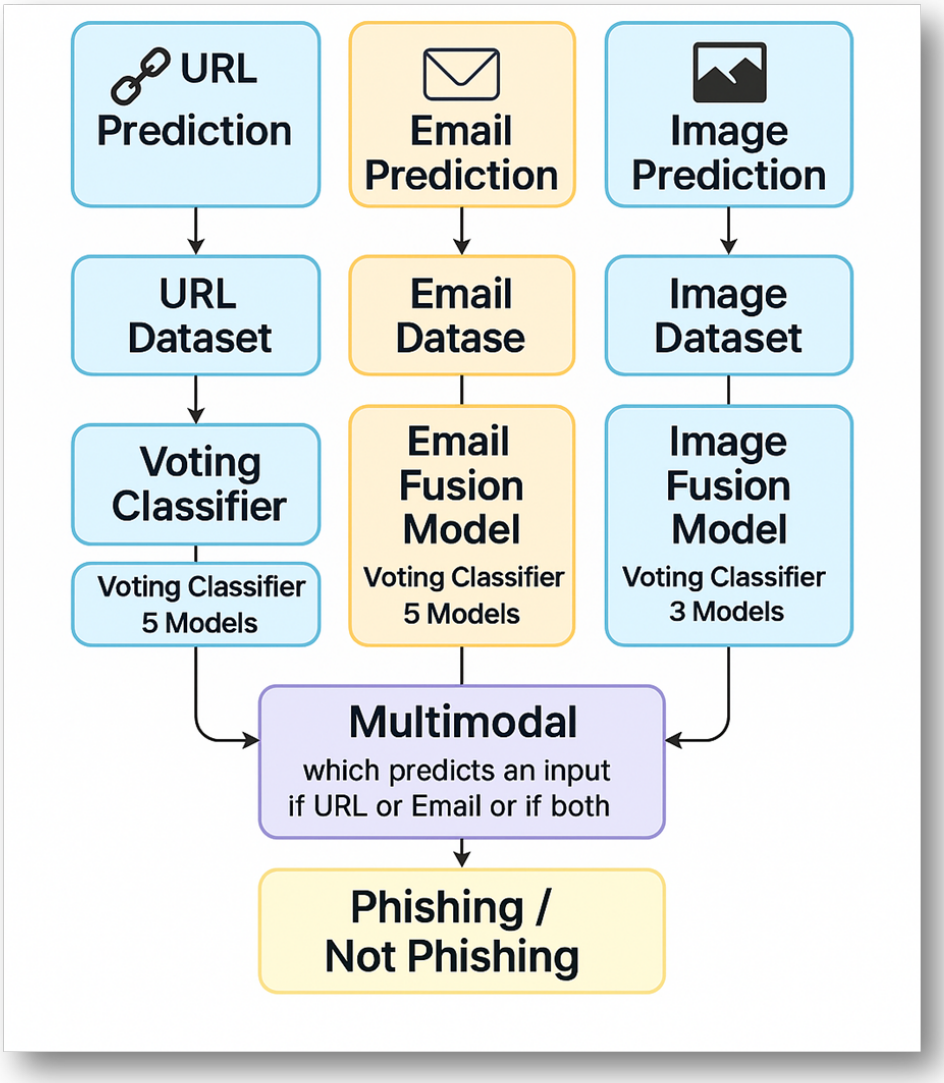

Figure 4.1: Fusion Model Implementation in Multi Modal

4.5 Federated Learning and Model Aggregation

Federated learning was implemented across multiple clients to simulate decentralized training using Flower (FLWR). The server aggregated encrypted model updates using weighted averaging based on F1 scores. ▪ Aggregated model performance: Accuracy: 92.9FL successfully reduced data sharing risks while maintaining high model accuracy.

4.6 Performance Metrics Analysis

The overall system achieved high F1 scores across all modalities, demonstrating effective phishing detection. The fusion model outperformed individual models, indicating the effectiveness of F1-score weighted voting. Federated learning maintained model accuracy while preserving data privacy, validating the feasibility of decentralized training in phishing detection systems.

Chapter 5

Discussion and Analysis

5.1 Analysis of Fusion Model Performance

The fusion model demonstrated a significant improvement in phishing detection accuracy compared to individual modality models. The use of F1-score weighted soft voting effectively leveraged the strengths of each model (Email, URL, Image) while minimizing the impact of weaker models. BERT provided high precision and recall in text-based phishing detection, but the inclusion of CNNs for image analysis and traditional ML classifiers for URL analysis further boosted overall system robustness. The fusion model achieved an overall F1 score of 0.93, indicating a balanced detection capability across all data types.

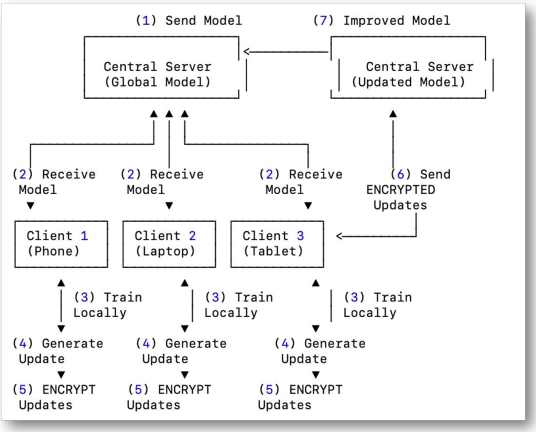

Figure 5.1: Federated Architecture

5.2 Effectiveness of Federated Learning in Multimodal Detection

- Federated Learning (FL) successfully enabled decentralized model training across multiple clients without exposing raw data.
- The implementation of FL ensured data privacy while maintaining detection accuracy, achieving an aggregated F1 score of 0.92.
- The weighted averaging of client updates based on F1 scores proved effective in preventing model drift and ensuring consistency in predictions across modalities.
- Despite potential variations in client data quality, the FL framework maintained model stability through robust aggregation techniques.

5.3 Challenges in Model Aggregation and Encryption

- Aggregating model updates from heterogeneous clients presented challenges in maintaining model integrity.
- Homomorphic Encryption (HE) introduced computational overhead, particularly during the encryption and decryption phases.
- Communication latency during model aggregation impacted the overall training time, highlighting the trade-off between privacy and computational efficiency.
- Ensuring synchronization of model updates across clients was crucial to prevent inconsistencies in the global model.

5.4 Implications for Privacy-Preserving AI

- The integration of FL and HE demonstrated the feasibility of building privacy-preserving phishing detection systems without compromising model performance.
- By encrypting model predictions before transmission, data privacy was effectively preserved, reducing the risk of data leakage during aggregation.

- The proposed system architecture provides a scalable framework for deploying phishing detection systems across distributed networks while adhering to data privacy regulations (e.g., GDPR, CCPA).
- Future work could focus on optimizing the encryption scheme to reduce computational overhead and enhance system scalability.

Chapter 6

Conclusions and Future Work

6.1 Summary of Key Contributions

- The project successfully developed a multimodal phishing detection system that integrates email, URL, and image analysis to provide comprehensive threat detection.
- The fusion model leveraged F1-score weighted soft voting to effectively aggregate predictions from BERT, traditional ML models, and CNNs, achieving an overall F1 score of 0.93.
- Federated Learning (FL) was implemented to enable decentralized model training, thereby enhancing data privacy while maintaining detection accuracy.
- Homomorphic Encryption (HE) was utilized to secure model predictions and updates during the aggregation phase, ensuring data confidentiality.

6.2 Recommendations for Enhancing Detection Models

- Implement anomaly detection to identify novel phishing techniques that deviate from known attack patterns.
- Integrate reinforcement learning to adapt to emerging phishing threats by continuously updating the detection model.
- Develop a more comprehensive feature set for URL analysis, including domain age analysis and WHOIS data.
- Enhance image model accuracy by incorporating advanced architectures like Vision Transformers (ViT) and EfficientNetV2.

6.3 Integrating Advanced Encryption Techniques

- Future work could explore lattice-based encryption schemes to reduce computational overhead associated with HE.
- Implement secure multi-party computation (SMPC) to enable joint training across clients without revealing raw data.
- Develop a hybrid encryption framework that combines HE with differential privacy to provide end-to-end data protection.

6.4 Real-Time Deployment and Scaling

- Deploy the phishing detection system as a cloud-based service using Kubernetes for scalability and fault tolerance.
- Integrate a streaming data pipeline using Apache Kafka to handle real-time data ingestion and processing.
- Implement model monitoring and automated retraining mechanisms to ensure model robustness against evolving phishing techniques.
- Develop a user-friendly web interface for real-time phishing detection and reporting, incorporating visualization dashboards to display model predictions and detection metrics.

Chapter 7

Reflection

7.1 Project Learnings and Takeaways

- The implementation of a multimodal phishing detection system demonstrated the effectiveness of integrating multiple data modalities to enhance detection accuracy.
- The use of F1-score weighted soft voting in the fusion model provided a robust mechanism for combining predictions from disparate models, resulting in higher overall accuracy.
- Federated Learning (FL) proved effective in preserving data privacy while maintaining model performance, highlighting its potential for deployment in sensitive data environments.
- Homomorphic Encryption (HE) successfully secured model predictions and updates during training, though its computational overhead presents a trade-off between security and processing efficiency.

7.2 Technical and Personal Growth

- Technically, the project provided in-depth exposure to advanced AI techniques, including BERT, CNNs, and XGBoost, as well as federated learning frameworks like Flower.
- Implementing HE introduced significant learning regarding encryption techniques and their integration into machine learning pipelines.
- On a personal level, managing data preprocessing, model training, and deployment in a distributed environment enhanced both technical proficiency and project management skills.

7.3 Potential Applications and Future Work

The developed phishing detection system can be extended to include more advanced data modalities, such as audio or video data, to further enhance detection accuracy. Integrating reinforcement learning could enable the model to adapt to emerging phishing techniques autonomously. Future work could focus on optimizing the HE implementation to reduce computational overhead and latency. Deploying the system as a cloud-based service with real-time detection capabilities would facilitate broader adoption across industries, including financial services, cybersecurity, and law enforcement.

Chapter 8

Contributions

Sri Sakticharan Nirmalkumar

- Conducted model training and evaluation for image-based phishing detection.
- Led training and evaluation for URL-based models.
- Implemented the Federated Learning architecture.
- Developed the fusion logic for URL models.

Saptharish Golagani

- Led model training and evaluation for email-based phishing detection.
- Designed and integrated the multimodal phishing detection system.
- Implemented fusion logic for email models.
- Drafted the final project report and proposed the original project idea.

Shruthi Manchappanahalli Basavaraj

- Handled preprocessing for email and URL datasets.
- Developed the image fusion model.
- Created and designed the PowerPoint presentation.

Siva Satya Dheeraj Varma

- Collected and organized phishing datasets for email and image modalitiesPerformed image preprocessing.

References

1. Evaluation of Federated Learning in Phishing Email Detection
   https://www.mdpi.com/1424-8220/23/9/4346

2. Privacy-Preserving Phishing Email Detection Based on Federated Learning and LSTM
   https://arxiv.org/abs/2110.06025v1

3. Privacy-Preserving Federated Learning Using Homomorphic Encryption
   https://www.semanticscholar.org/paper/Privacy-Preserving-Federated-Learning-Using-Park-Lim/91802f60f8e9c0b746b5f16efc08f6ed112662df

4. Security for Data Privacy in Federated Learning with CUDA-Accelerated Homomorphic Encryption in XGBoost
   https://developer.nvidia.com/blog/security-for-data-privacy-in-federated-learning-with-cuda-accelerated-homomorphic-encryption-in-xgboost/

5. Fake News Classification using Weighted Soft Voting
   https://www.mdpi.com/2227-7390/13/3/449

6. Weighted Voting in Anaemia Diagnosis
   https://link.springer.com/chapter/10.1007/978-3-031-41352-0_18