NETWORK INTRUSION DETECTION SYSTEM

Project Reference No.: *49S_BE_2129*

College : Bangalore Technological Institute, Bengaluru
Branch : Department of Computer Science and Engineering
Guide(s) : Prof. Shylaja D.N
Dr. Tamilarasan S
Student(s) : Ms. Raksha Kishor Kindalkar
Mr. Pranay Bandaru
Mr. Prajwal Hoysal Kumar P K
Mr. Malik Rehan

Keywords:

Network Security, Phishing Detection, Machine Learning, Gradient Boosting, Intrusion Detection System, Cybersecurity, Feature Extraction, URL Analysis, Real-time Detection

Introduction:

1. The rapid growth of internet-based services such as online banking, e-commerce, and cloud platforms has significantly increased the risk of cyber threats. Among these threats, phishing attacks are one of the most common and dangerous forms of cybercrime, where attackers deceive users into revealing sensitive information like passwords and financial data. Traditional security mechanisms such as firewalls and blacklist-based detection systems are no longer sufficient to detect modern phishing attacks, especially zero-day attacks.
2. Network Intrusion Detection Systems (NIDS) have evolved to address these challenges by analyzing network traffic and identifying malicious activities. However, conventional NIDS rely heavily on signature-based techniques, which fail to detect unknown threats. Machine learning techniques offer a promising solution by learning patterns from data and identifying anomalies.
3. This project focuses on developing a machine learning-based phishing detection system using a Gradient Boosting classifier. The system extracts features from URLs, domain information, and webpage content to accurately classify websites as legitimate or phishing. The proposed solution aims to provide real-time detection, high accuracy, and reduced false positives, thereby enhancing cybersecurity and protecting users from online threats.

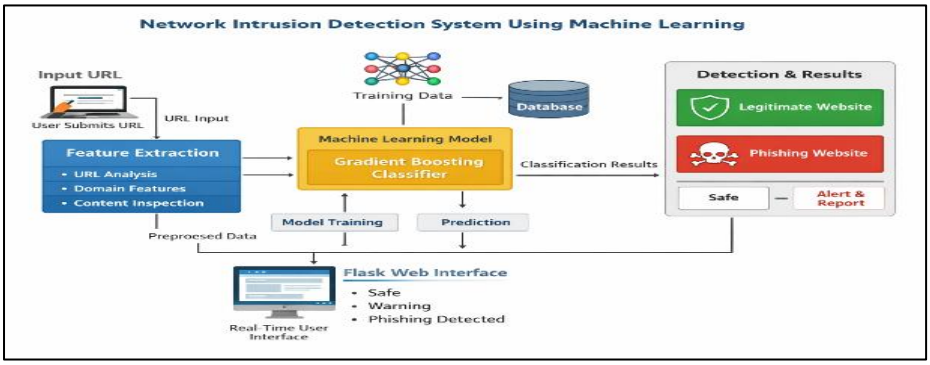

Figure 1: General Workflow of the Proposed NIDS Model

Objectives:

1. To analyse URL structures, domain attributes, and webpage features to distinguish between legitimate and malicious sites
2. To implement a machine learning model using the Gradient Boosting technique for accurate classification.
3. To achieve high detection accuracy while minimizing false positives and false negatives
4. To develop a real-time system for identifying malicious URLs through a web-based interface.
5. To enhance cybersecurity by providing an efficient and scalable detection system.

Methodology:

1. Dataset Collection:
   The methodology begins with collecting a comprehensive dataset of both phishing and legitimate URLs from publicly available and reliable sources. The dataset is balanced to ensure equal representation, reducing bias and improving model performance.
2. Data Preprocessing:
   The collected dataset is cleaned by handling missing values, removing duplicate entries, and eliminating irrelevant data. Feature values are normalized to maintain consistency and improve the quality of input data.
3. Feature Extraction:
   Relevant features are extracted from URLs and webpages, including URL-based features (length, special characters, IP usage), domain-based features (domain age, registration details), and content-based features (HTML structure and JavaScript behaviour).
4. Feature Selection and Optimization**:**
   Unnecessary and redundant features are removed to reduce complexity. A refined set of meaningful features is selected to improve classification accuracy and make the model more efficient.
5. Model Development:
   A machine learning model is developed using the Gradient Boosting technique, which combines multiple weak learners to form a strong predictive model capable of identifying complex phishing patterns.
6. Model Training and Validation:
   The dataset is divided into training and testing sets. The model is trained on the training data and validated using testing data along with cross-validation techniques to ensure good generalization and avoid overfitting.
7. System Implementation:
   A web-based application is developed using Flask to provide a user-friendly interface. Users can input URLs, and the system performs real-time feature extraction and classification.

8. Prediction and Result Display:
   When a user inputs a URL, the system performs real-time feature extraction and classification. The result is displayed instantly, indicating whether the URL is safe or phishing along with a confidence score.
9. Testing and Performance Evaluation:
   The system is tested using real-world URLs to ensure reliability and robustness. Performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to validate the effectiveness of the model.
10. Deployment Consideration:
    The final system is evaluated for deployment in real-time environments such as web browsers or network security systems. Scalability and usability are analyzed to ensure practical implementation.

Results and Conclusion:

1. In conclusion, the developed system successfully detects phishing websites with high accuracy and reliability. The Gradient Boosting model demonstrated strong performance in distinguishing between legitimate and malicious URLs.
2. Key observations include:
   - High detection accuracy (above 90%)
   - Reduced false-positive rates compared to traditional systems
   - Efficient real-time processing of user inputs
   - Effective identification of zero-day phishing attacks
3. Simulation results confirmed that the system can handle various types of URLs, including complex and obfuscated phishing links. The user interface provides clear outputs, including safety scores and prediction confidence. The project concludes that machine learning-based intrusion detection systems are more effective than traditional rule-based approaches. The proposed system provides a scalable, efficient, and intelligent solution for enhancing cybersecurity.

Project Outcome & Industry Relevance:

This project contributes significantly to the field of cybersecurity by providing an intelligent phishing detection system. The developed model can be integrated into:

- Web browsers for real-time URL scanning
- Email filtering systems to detect malicious links
- Enterprise security systems for network monitoring
- Financial institutions for fraud prevention

The system is highly relevant to industries such as:

- Banking and Finance E-commerce
- IT Security Services
- Cloud Computing Platforms

By reducing phishing attacks, the system helps organizations minimize financial losses and improve data security. It also enhances user awareness and trust in online systems.

Working Model vs. Simulation/Study:

The project is developed as a working model with a real-time web application for phishing detection. It allows users to input URLs and performs live feature extraction and

classification using a trained machine learning model. Unlike simulation-only studies, the system works dynamically on real-world inputs. In addition, simulation and testing were carried out using datasets to evaluate accuracy and performance. The integration of both real-time implementation and simulation ensures reliability and practical usability.

Project Outcomes and Learnings:

1. Gained knowledge in machine learning algorithms and model optimization.
2. Learned feature extraction techniques for cybersecurity applications.
3. Developed skills in Python, Flask, and database management.
4. Understood real-time system integration and deployment.
5. Improved problem-solving and research skills in cybersecurity domain.

References:

1. [1] F. Guo, et al., “Information Security Network Intrusion Detection System Based on Machine Learning,” International Conference on Data Science and Network Security (ICDSNS), 2024, pp. 1–6.
2. [2] P. M. Kumar, et al., “Enhanced Network Intrusion Detection System Using PCGSO-Optimized BI-GRU Model in AI-Driven Cybersecurity,” *IEEE International Conference on Artificial Intelligence in Cybersecurity (ICAIC)*, 2024.
3. [3] N. D. Patel, et al., “Detection of Intrusions Using Support Vector Machines and Deep Neural Networks,” *International Conference on Recent Innovations in Technology and Optimization (ICRITO)*, 2022, pp. 120–125
4. [4] S. Sridevi, “Network Intrusion Detection System Using Supervised Learning Based Voting Classifier,” *International Conference on IoT and Smart Systems*, 2022.
5. [5] J. Abraham, et al., “Intrusion Detection and Prevention in Networks Using Machine Learning and Deep Learning Approaches,” *IEEE Conference*, 2021.

Future Scope:

1. The future scope of this project includes integrating advanced deep learning models such as LSTM, GRU, and CNN to improve detection of complex phishing attacks. Real-time stream processing using tools like Apache Kafka or Spark Streaming can be implemented for faster detection in large-scale networks.
2. The system can be enhanced by incorporating multi-layer security data such as firewall logs, DNS records, and endpoint monitoring. Adaptive learning techniques can be introduced to allow the model to update dynamically with new attack patterns.
3. Natural Language Processing (NLP) can be used to analyze email content and webpage text for improved phishing detection. Deployment as a cloud-based service will enable scalability and accessibility for organizations. Visualization dashboards can be added for better monitoring and analysis.
4. Automated response mechanisms can be implemented to block malicious URLs instantly Integration with blockchain technology can ensure secure logging and prevent tampering of intrusion data.

   These enhancements will transform the system into a comprehensive, enterprise-level cybersecurity solution