**Name: Aliyan Ahmed Cheema**

Reg # FA22-BCE-028

**AI-Driven Cybersecurity: Anomaly Detection Using Isolation Forest Based on the CyberSentinel Framework**

**Abstract:**

As cyber threats continue to evolve with increasing complexity, traditional rule-based detection systems have proven inadequate in identifying zero-day attacks and other emergent threats. In response, AI-driven approaches are being integrated into cybersecurity systems. This report focuses on a replication and practical implementation inspired by the research paper titled “CyberSentinel: An Emergent Threat Detection System for AI Security” by Dr. Krti Tallam. Using publicly available SSH login data, we simulate key features described in the CyberSentinel paper and build a lightweight anomaly detection model using the Isolation Forest algorithm. The experiment showcases how AI can effectively detect outliers that may indicate cybersecurity threats. Evaluation metrics such as accuracy, precision, recall, and F1 score are calculated, and anomaly score distributions are visualized to assess model performance.

1. **Introduction**

The growing reliance on AI systems in critical applications has introduced novel cybersecurity vulnerabilities. Traditional static detection methods, which rely heavily on known signatures or manual rules, fall short in detecting unknown or evolving threats. Anomaly detection, particularly through unsupervised machine learning techniques, offers a viable alternative. The CyberSentinel framework proposes a unified agent that utilizes multiple modules, including SSH brute-force detection, phishing detection, and an emergent threat detection engine. The emergent threat detection component specifically relies on unsupervised learning models like Isolation Forest to identify anomalous behavior in login patterns and system usage. In this report, we implement a simplified version of the ETD module using real-world-like SSH login data.

1. **Literature Review**

Traditional cybersecurity solutions have focused on identifying known threats using signature databases and heuristic rules. However, these approaches are limited in scope, particularly in cloud and AI-native environments where behaviors evolve rapidly. Recent research has focused on anomaly-based methods using unsupervised learning algorithms. Isolation Forest, introduced by Liu et al. in 2008, is particularly efficient in high-dimensional settings and has been widely used in cybersecurity for detecting intrusions, fraud, and spam. The CyberSentinel paper advances this line of work by combining different detection methods under a unified framework and enhancing system flexibility through modular design. By automatically retraining its models, CyberSentinel adapts to new data trends, making it more resilient against sophisticated threats.

**3. Methodology**

**3.1 Dataset Preparation**

We used a publicly available dataset containing SSH login events with features such as ip\_failure, not\_valid\_count, td (time difference between login attempts), and class labels. Since the original CyberSentinel dataset was unavailable, we simulated key features such as: - ip\_numeric: Randomly assigned integer identifiers for IP addresses - geo\_distance: Simulated geolocation distances from a hypothetical origin point - hour: Extracted from UNIX timestamp or randomly simulated where timestamps were absent

**3.2 Feature Selection**

We selected six features for our anomaly detection model: - hour - ip\_numeric - geo\_distance - ip\_failure - td - not\_valid\_count

**3.3 Model and Training**

We used the Isolation Forest algorithm with 100 estimators and 10% contamination rate. The dataset was split 70% for training and 30% for testing. Isolation Forest assigns anomaly scores based on how isolated a data point is within the feature space. Predictions were classified as 1 for anomaly and 0 for normal.

**4. Results and Analysis**

**4.1 Anomaly Score Distribution**

Histograms of anomaly scores showed that most login attempts clustered around high-density regions, with outliers easily distinguishable. Both training and test sets showed similar score distributions, indicating model stability.

**4.2 Classification Metrics**

Based on ground truth labels in the dataset, the following metrics were obtained: - Accuracy: 0.93 - Precision: 0.79 - Recall: 0.88 - F1 Score: 0.83

These results demonstrate that even with simulated features, the Isolation Forest model was able to detect anomalous behavior with high accuracy and reliability.

**4.3 Anomaly Count Visualization**

A bar chart was used to visualize the number of normal vs. anomalous predictions for both train and test datasets. The ratio of anomalies remained consistent, aligning well with the set contamination rate and ground truth labels.

**5. Discussion**

The experiment replicates the core ideas of the CyberSentinel ETD module. Despite the absence of the original dataset, simulated features based on the paper’s descriptions allowed for a credible approximation of the detection mechanism. The high accuracy and F1 score affirm the practicality of unsupervised models in real-world cybersecurity contexts. This approach is particularly useful for detecting emerging threats that don’t match known attack patterns. However, the model is sensitive to feature quality, and simulated data may not fully capture real-world variability. Future efforts should focus on accessing richer datasets and incorporating additional real-time features like login source, user-agent, and behavior history.

**6. Conclusion and Future Work**

This project demonstrates the feasibility and effectiveness of AI-driven cybersecurity systems like CyberSentinel. Using Isolation Forest, we achieved high detection performance on a realistic SSH login dataset. The modular nature of the model supports easy expansion to include additional threat detection modules. In the future, integration with real-time log data and the use of explainable AI techniques could enhance transparency and trust. Moreover, deploying the model in a live environment would provide better insight into its resilience and adaptability.

**References** - Tallam, Krti. “CyberSentinel: An Emergent Threat Detection System for AI Security.” arXiv preprint arXiv:2502.14966 (2025). - Liu, Fei Tony, Ting, Kai Ming, and Zhou, Zhi-Hua. “Isolation forest.” 2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008. - Osama C., “SSH Logs with Attack Classification,” Kaggle, https://www.kaggle.com/datasets/osamac/ssh-logs-with-attack-classification