Research Paper Sections

# I. Background of the Study

With the escalating complexity and frequency of cyberattacks, traditional security mechanisms often fall short, especially in host-based environments. This necessitates the development of advanced Host-Based Intrusion Detection Systems (HIDS) capable of detecting subtle, sophisticated, and dynamic threats in real-time. The research utilizes system call sequence analysis from three key datasets—ADFA-LD, ADFA-WD, and ADFA-WD:SAA—to enhance detection accuracy and responsiveness using hybrid machine learning and deep learning techniques. While previous research has explored various models and feature engineering methods, real-time processing capabilities and effective detection of stealth attacks remain ongoing challenges, forming the core motivation behind this study.

# II. Statement of the Problem

Despite numerous advancements in HIDS, current systems still struggle to detect stealthy and complex attacks in real-time, particularly under the constraints of class imbalance and system noise. The research aims to address the limitations of existing HIDS in accurately detecting sophisticated and stealth-based intrusions across Linux and Windows environments using system call analysis.

# III. Objectives of the Study

The main objectives of this research are:

1. To enhance intrusion detection accuracy in HIDS by leveraging hybrid machine learning and deep learning models.

2. To develop efficient and robust feature extraction techniques for system call analysis.

3. To address data imbalance issues in security datasets to improve detection rates of rare attacks.

4. To evaluate the proposed models across the ADFA-LD, ADFA-WD, and ADFA-WD:SAA datasets.

5. To ensure real-time performance in host-based intrusion detection.

# IV. Hypotheses (if applicable)

H1: A hybrid model combining traditional machine learning and deep learning will achieve higher detection accuracy than standalone models.

H2: Advanced feature extraction techniques (like TF-IDF+SVD, CNN-based learning) significantly improve model performance.

H3: Implementing data imbalance handling techniques (e.g., loss functions or sampling) will improve the detection of low-frequency attacks.

# V. Significance of the Study

This study contributes a practical and scalable HIDS framework capable of real-time intrusion detection across Linux and Windows platforms. The work addresses critical gaps in current methods by integrating deep learning and ensemble learning strategies to detect stealth attacks and reduce false positives. The research offers practical deployment insights, thereby benefitting cybersecurity practitioners and organizations looking to fortify their endpoint protection systems.

# VI. Scope and Limitations of the Study

Scope:

- Analysis focuses on system call-based intrusion detection using ADFA datasets (ADFA-LD, ADFA-WD, ADFA-WD:SAA).

- Evaluation is limited to machine learning and deep learning models for anomaly and pattern detection.

- Experiments consider early detection, feature engineering, and model accuracy/efficiency metrics.

Limitations:

- Limited to system call logs from controlled environments (may not capture all real-world attack variants).

- Performance may degrade with extremely long sequences or incomplete traces.

- Generalizability across unseen datasets or newer operating systems remains a challenge.