**Anomaly Defense System Using Intrusion**

**Detection and Anomaly Identification**

Urmishikha Dash   
*BTech AI*  
*Manipal Institute of Tech  
Bengaluru*  
[urmishikha.mitblr2022@learner.manipal.edu](mailto:urmishikha.mitblr2022@learner.manipal.edu)

*Abstract*— The cyber threats are becoming ever more sophisticated, necessitating the development of ever more advanced systems to identify and respond to anomalies in network traffic. This paper proposes an AI-based anomaly defense system that merges intrusion detection and anomaly identification in order to improve network security. It relies on machine learning algorithms, real-time monitoring, and automatic reaction; it can thus detect and effectively neutralize possible cyberattacks.

# Introduction

The rapid development of digital networks that can be penetrated now expose cyber threats comprising unauthorized access, denial of service (DoS), and data breaches. Signature-based intrusion detection systems (IDS) fail against new threats. This paper puts forth anomalous detection systems using machine learning and real-time monitoring techniques to dynamically detect and prevent cyber threats.

# LITERATURE REVIEW

NLP has evolved from rule-based systems to deep learning models capable of sophisticated text understanding. OCR technology, particularly Tesseract, plays a pivotal role in extracting text from images, contributing to NLP's growing application in document processing. Models like Gemini LLM leverage extensive training data to provide contextually relevant answers. Prior studies emphasize the effectiveness of combining OCR and LLMs in document management, showing that NLP-based tools can dramatically reduce manual effort and improve data accessibility.

# METHODOLOGY

The proposed system comprises the following components:

Network traffic data includes tools like Wireshark and Snort. Packet characteristics like source IP, type of protocol, and packet processing rate.

Unsupervised learning models (K-Means and DBSCAN) to identify deviations from a normal pattern. Supervised models (Random Forest, Logistic Regression) to classify various types of attacks.

Suspicious IPs are blocked by automatic scripts and alerts raised. Streamlit interface enables easy uploads and queries, optimizing document management for diverse formats and enhancing data accessibility and productivity.

# OBJECTIVES

* **Automate Document Processing**: Develop an efficient pipeline to automate text extraction from various document formats, including PDFs, images, and handwritten notes.
* **Build a Centralized Knowledge Base**: Organize extracted data in a structured, searchable format for easy retrieval.
* **Enhance User Accessibility**: Create a user-friendly interface with Streamlit for simple, intuitive document uploads and querying.
* **Provide Contextual Responses**: Leverage the Gemini LLM to deliver accurate, context-aware answers, improving information retrieval and decision-making.

# ADVANTAGES

* **Efficiency**: Automated text extraction saves time, especially when handling large volumes of data.
* **High Detection Accuracy**: Machine learning models enhance the capability of the system in identifying unknown threats.
* **Real-time Monitoring**: Continuous network monitoring enables timely actions for security threats.
* **Automated Threat Mitigation**: The system can automatically take proactive security measures, such as blocking malicious IP addresses.
* **Scalability**: The architecture supports real-time data ingestion and classification, providing it with the capability to be extended across large-scale networks.

# DISADVANTAGES

* **Dependence on OCR Quality**: Accuracy of text extraction may vary based on the quality of input images and OCR performance.
* **Complex Queries Limitation**: While the LLM performs well with structured questions, ambiguous or overly complex queries could challenge its accuracy.
* S**calability constraints**: The processing pipeline may require optimization to handle very large or simultaneous data uploads efficiently.

# FUTURE ENHANCEMENTS

* **Reinforcement Learning**: Enhancing Adaptive Threat Detection and Mitigation Strategies.
* **Cloud-based Integration**: Deploying the system on cloud security platforms to manage an excessive amount of data.
* **Advanced Feature Engineering**: Incorporating more sophisticated network traffic measurement features for better detection.
* **Blockchain for Security Logs**: Deploying an immutable logging solution based on blockchain technology to prevent log tampering.
* **Integration with SIEM Systems**: Linking with Security Information and Event Management platforms for enterprise-wide monitoring.

# RESULTS

The model is evaluated using metrics such as precision, recall, and F1-score. Testing on NSL-KDD dataset shows high accuracy in detecting DoS, brute force, and SQL injection attacks. Real-time performance is assessed by deploying the system in a simulated network environment, demonstrating its capability to trigger alerts and automate responses effectively.

# CONCLUSION

This paper presents a machine learning-powered cybersecurity solution for real-time anomaly detection and intrusion response. Future improvements will include reinforcement learning for adaptive threat mitigation and integration with cloud security platforms.

##### References

1. Ahmed, Mohiuddin, Abdun Naser Mahmood, and Jiankun Hu. "A survey of network anomaly detection techniques." *Journal of Network and Computer Applications* 60 (2016): 19-31.
2. Pang, Guansong, et al. "Deep learning for anomaly detection: A review." *ACM computing surveys (CSUR)* 54.2 (2021): 1-38.
3. Goldstein, Markus, and Seiichi Uchida. "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data." *PloS one* 11.4 (2016): e0152173.
4. Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation-based anomaly detection." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 6.1 (2012): 1-39.
5. Patcha, Animesh, and Jung-Min Park. "An overview of anomaly detection techniques: Existing solutions and latest technological trends." *Computer networks* 51.12 (2007): 3448-3470.