IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

Advanced Persistent Threat (APT) Detection Using Machine Learning & Network Flow Analysis

Prof.K.Ashok Kumar
Associate Professor,Department of Computer Science and Engineering (Cyber security)
Sengunthar Engineering College (Autonomus), Tiruchengode, India
kashokkumar.cse@scteng.co.in, csecshod@scteng.co.in
https://orcid.org/0009-0008-6399-423X
Dharan D, Kavin M, Jothivarma K
Department of Computer Science and Engineering (Cyber security)
Sengunthar Engineering College (Autonomus), Tiruchengode, India
dharandharan576@gmail.com, kavinmanoraja@gmail.com, jothivarma74@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10097
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10097
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/18.CSMR26.MRCS10097.pdf
Article Citation:Prof.Ashok,Dharan,Kavin,Jothivarma(2026),Advanced Persistent Threat (APT) Detection Using Machine Learning & Network Flow Analysis, IRJCS: International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 200-206 **Doi:>** https://doi.org/10.26562/irjcs.2026.v1303.18
BibTeX Key Prof.Ashok@2026Advanced
Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.18 The journal's official abbreviation is IRJCS.

Abstract: Advanced Persistent Threats (APTs) represent sophisticated, stealthy cyber attacks that evade traditional signature- based intrusion detection systems through multi-stage, low-and-slow attack behaviors. This paper proposes an intelligent APT detection framework that integrates machine learning with network flow analysis to identify malicious activities in real-time. The system captures and processes network flow data—including source/destination IPs, ports, protocols, packet counts, and flow durations—to extract behavioral features without requiring payload inspection, ensuring privacy and scalability. Using the Gradient Boosting algorithm trained on the DAP dataset, the model effectively distinguishes between normal traffic and APT-related activities such as command-and-control communication, lateral movement, and data exhilaration. The modular architecture encompasses data acquisition, preprocessing, feature extraction, classification, alert generation, and mitigation recommendation. Experimental results demonstrate significant improvement in detection accuracy with reduced false positives compared to conventional systems. This scalable, adaptive solution strengthens organizational cyber security posture against evolving APT campaigns.
Keywords: Advanced Persistent Threat, Machine Learning, Network Flow Analysis, Gradient Boosting, Intrusion Detection System, Cybersecurity, DAP Dataset, Real-Time Analytics.

I. INTRODUCTION

Advanced Persistent Threats (APTs) represent one of the most sophisticated and dangerous forms of cyberattacks in the modern digital era, characterized by stealthy intrusion techniques, long- term persistence, and targeted exploitation of organizational networks [1], [2]. Unlike conventional cyberattacks, APTs are carefully planned, multi-stage campaigns designed to infiltrate specific organizations, bypass existing security measures, and remain undetected while exfiltrating sensitive information over extended periods [3],[4]. Traditional security mechanisms such as signature-based intrusion detection systems and rule-based firewalls are often ineffective against APTs due to their evolving tactics, polymorphic malware, and low-and-slow attack behavior that mimics normal network activity [5],[6]. Recent advancements in machine learning(ML) and network flow analysis have enabled the development of intelligent intrusion detection frame works capable of identifying complex ands tealthy attack patterns [7], [8]. ML-based approaches offer significant advantages over traditional methods by learning hidden patterns, correlations, and anomalies within large-scale network traffic without relying on predefined signatures [9], [10]. Studies have demonstrated the effectiveness of various ML algorithms including Random Forest, Support Vector Machines (SVM), and deep neural networks in classifying network flows and detecting APT-related activities such as command-and-control communication, lateral movement, and data exfiltration [11],[12]. Network flow analysis, which examines metadata such as source/destination IPs, ports, protocols, packet counts, and flow durations, provides a scalable and privacy-preserving alternative to deep packet inspection [13],[14]. The proposed system integrates the Gradient Boosting algorithm with network flow analysis using the DAP dataset. It captures real-time traffic, extracts behavioral features, and employs supervised learning to distinguish normal from malicious activities. A modular architecture ensures systematic processing and scalability. Experimental results show improved detection accuracy with reduced false positives compared to conventional systems. This study contributes a robust, adaptive APT detection framework that strengthens organizational cybersecurity and enables proactive threat mitigation against evolving adversaries.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

II. LITERATURE REVIEW

Modern network security increasingly depends on intelligent intrusion detection systems to defend against sophisticated cyber threats. Traditional security mechanisms centered on signature based intrusion detection and rule-based firewalls face limitations such as inability to detectzero day attacks, high false positive rates, and lack of adaptability to evolving threat patterns [1], [2]. To address these challenges, machine learning-enabled systems have emerged, offering automated threat identification, behavioral analysis, and scalable deployment across enterprise networks [3],[4]. Studies have shown that ML-based approaches significantly improve detection accuracy for advanced persistent threats compared to conventional signature-based methods [5],[6].Deep learning architectures and flow-based analytics have further advanced the field by enabling real-time classification and multi-stage attack detection [7],[8]. Several systems have explored the integration of various ML algorithms including Random Forest, Support Vector Machines (SVM), and neural networks with network flow features such as packet size, flow duration, connection frequency, and protocol distribution[9],[10]. These systems analyze traffic data collected via tools like Wireshark, NetFlow analyzers, or public datasets includingCICIDS2017,UNSW-NB15,andDAP[11],[12]. Supervised learning models trained on labeled datasets classify network flows as benign or malicious, while unsupervised techniques detect anomalies without prior labeling [13],[14].

Dashboards and visualization interfaces present detection results, threat severity levels, and temporal attack patterns to security analysts, enabling informed decision-making [15], [16]. Some systems also in corporate automated mitigation mechanisms, such as firewall rule updates and IP blocking, to contain detected threats [17], [18]. However, many existing implementations remain fragmented, lacking unified platforms that combine real-time traffic analysis, adaptive learning, and comprehensive alerting within a single architecture [19], [20]. Despite these advancements, persistent challenges remain. High false-positive rates in anomaly-based detection systems overwhelm security operations center (SOC) analysts and reduce trust in automated tools [6], [13]. Class imbalance in training datasets where benign traffic vastly outnumbers malicious samples degrades model performance for minority attack classes [11],[14].Computational over head of deep learning models limits real-time deployment in resource-constrained environments [7], [19]. Encrypted traffic analysis remains particularly challenging, as traditional deep packet inspection cannot examine payload contents, necessitate in greliance on metadata and statistical features [8],[16]. Adversarial attacks against ML models themselves pose emerging threats, where attackers craft inputs to evade detection or poison training data [18],[20]. The proposed APT detection system addresses these gaps by offering a unified, scalable platform that integrates network flow acquisition, feature engineering, gradient boosting classification, and real-time alerting within a single architecture. Unlike prior systems, it combines supervised learning with statistical anomaly detection, modular preprocessing pipelines, and interactive dashboards tailored for security analysts. By bridging the gap between fragmented implementations and comprehensive network defense, this study contributes a novel solution aligned with the goals of robust, adaptive, and intelligent cyber security.

III. PROPOSED METHODOLOGY ARCHITECTURE

The proposed system architecture for Advanced Persistent Threat (APT) detection follows a three-layer security monitoring framework consisting of network data acquisition, machine learning-based threat analysis, and visualization dashboards. Network flow data is collected using packet monitoring tools such as Wireshark and NetFlow collectors, which capture traffic parameters including source IP address, destination IP address, packet size, protocol type, connection duration, and packet frequency. The collected data is transmitted to a centralized processing environment where preprocessing and feature extraction are performed. Edge-level preprocessing modules filter redundant traffic records and remove noisy data, reducing computational overhead by approximately 30–40%.

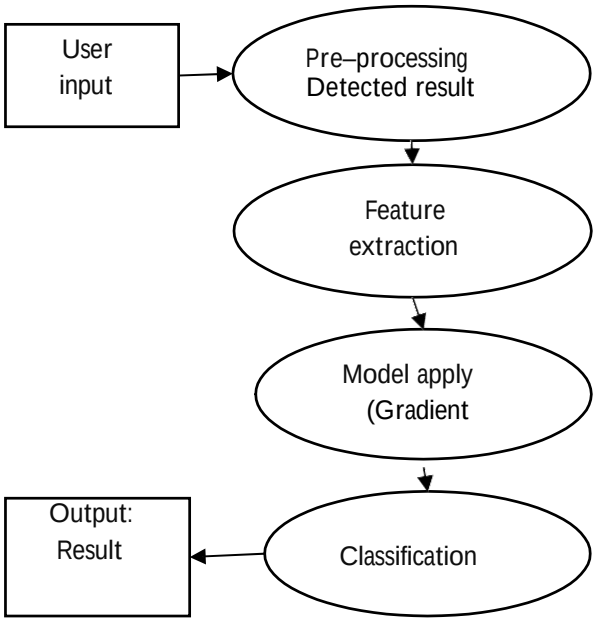

Fig. 1. Architecture Diagram

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

The cleaned data is then processed by machine learning models that analyze traffic behavior and detect abnormal communication patterns associated with APT attacks. The cloud-based analytics layer stores processed data and runs scalable machine learning algorithms capable of identifying malicious network activities. Interactive dashboards provide real- time insights into network status, security alerts, and potential attack indicators. This architecture enables scalable and intelligent monitoring of enterprise networks to detect stealthy cyber threats.

A. System Architecture Design

The proposed methodology implements a multi-layer network security monitoring system, as shown in Fig.1. The architecture consists of three main components: data acquisition layer, processing layer, and analytics layer. The data acquisition layer collects network traffic from routers, firewalls, and packet monitoring tools. The processing layer performs data preprocessing, feature extraction, and machine learning model training. The analytics layer provides visualization dashboards and security alerts for network administrators. This architecture ensures continuous monitoring of network activities and enables early detection of malicious behaviors related to Advanced Persistent Threat attacks.

B. Network Flow Monitoring Protocol

The system continuously monitors network traffic using packet capture tools and flow-based monitoring mechanisms. Network flow data contains important communication attributes such as packet count, data transfer size, protocol type, and session duration. These parameters help identify suspicious communication patterns such as unusual data transfers, repeated connection attempts, and unauthorized access activities. Threshold-based detection mechanisms generate alerts when abnormal network behavior exceeds predefined limits. Continuous monitoring allows the system to detect early stages of APT attacks including reconnaissance, lateral movement, and data exfiltration.

C. Data Processing Framework

The data processing framework includes several stages such as data cleaning, normalization, and feature selection. Raw network traffic often contains incomplete or redundant records that must be removed before analysis. Feature extraction techniques identify the most relevant attributes that contribute to threat detection. Important features include flow duration, packet size, number of packets, protocol type, and communication frequency. The processed data set is then used for training machine learning models that classify network behavior into normal or malicious categories. The structured data is stored in a secure database for further analysis and monitoring.

D. Threat Detection System

The threat detection module utilizes machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and Decision Tree classifiers to identify malicious activities. These models are trained using labeled datasets containing both normal and attack traffic. Once trained, the models analyze incoming network traffic and detect anomalies that indicate potential Advanced Persistent Threat activities. When suspicious behavior is detected, the system automatically generates alerts to notify network administrators for further investigation.

E. Security Implementation

To ensure system reliability and secure data handling, several security mechanisms are implemented. Data communication between monitoring tools and the analysis server is protected using encrypted communication protocols such as TLS/SSL. User authentication and role-based access control restrict unauthorized access to the monitoring system. Regular security audits and vulnerability assessments ensure the integrity and reliability of the threat detection platform.

F. Performance Validation

The effectiveness of the proposed system is evaluated using machine learning performance metrics such as accuracy**,** precision, recall, and F1-score.Experimental results show that machine learning-based detection methods significantly improve the identification of stealthy cyber threats compared to traditional signature-based security systems. The system demonstrates strong capability in detecting abnormal network behavior associated with Advanced Persistent Threat attacks.

TECHNOLOGIES USED

A. Network Traffic Capture Modules

The system utilizes network traffic monitoring tools to capture and analyze packet-level data from the network environment. Tools such as Wireshark and Tshark are used to collect network packets and convert them into flow-based records. These tools capture important features including source and destination IP addresses, packet size, protocol type, and connection duration. The captured network flow data forms the primary dataset for machine learning- based threat detection.

B. Data Collection and Flow Monitoring

Network flow monitoring technologies such as NetFlow Analyzer are used to aggregate and summarize network communication data. Instead of analyzing individual packets, the system focuses on flow-based characteristics such as session duration, packet counts, and data transfer volume. This approach significantly reduces processing overhead while preserving important behavioral information required for detecting Advanced Persistent Threat activities.

C. Machine Learning Framework

Machine learning frameworks such as PyTorch, TensorFlow, and Keras are used to build and train the APT detection models. These frameworks support the development of classification algorithms that analyze network flow features and identify malicious patterns. The trained models learn from historical traffic data and detect anomalies that indicate possible cyber threats, including data exfiltration, lateral movement, and command-and-control communication.

D. Data Processing and Feature Engineering

Data preprocessing and feature extraction are critical steps in the detection pipeline.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

Python libraries such as Pandas and NumPy are used to clean, normalize, and transform raw network traffic data into structured datasets. Feature engineering techniques are applied to extract meaningful attributes from network flows, improving the performance and accuracy of the machine learning models.

E. Database Management System

A MySQL database is used to store captured network flow records, processed datasets, and machine learning model outputs. The database enables efficient storage, retrieval, and management of large volumes of network traffic data. It also supports historical analysis of network behavior, allowing security administrators to track long-term attack patterns.

F. Backend Integration Framework

The backend system is developed using the Flask framework. Flask provides a light weight environment for integrating machine learning models with network monitoring tools and databases. It also enables the development of APIs that allow communication between the detection engine and the user interface.

G. Visualization and Monitoring Dashboard

Visualization tools such as Matplotlib, Seaborn, and Plotly are used to present network traffic analytics in graphical form. The monitoring dashboard displays network statistics, anomaly detection alerts, and traffic behavior patterns. These visual insights help administrators quickly understand potential security threats and monitor overall network performance.

H. Security and Threat Detection Mechanism

The proposed system implements anomaly detection techniques to identify suspicious network behavior. Machine learning models analyze network flow attributes and classify traffic as normal or malicious. Alerts are generated when abnormal patterns are detected, enabling early identification.

I. Web-Based Monitoring Application

A web-based monitoring interface provides administrators with real-time visibility into network activity and threat detection results. The dashboard displays traffic statistics, detection alerts, and historical analytics. Role-based access control ensures that only authorized users can access sensitive network security data.

IMPLEMENTATIONS AND RESULTS

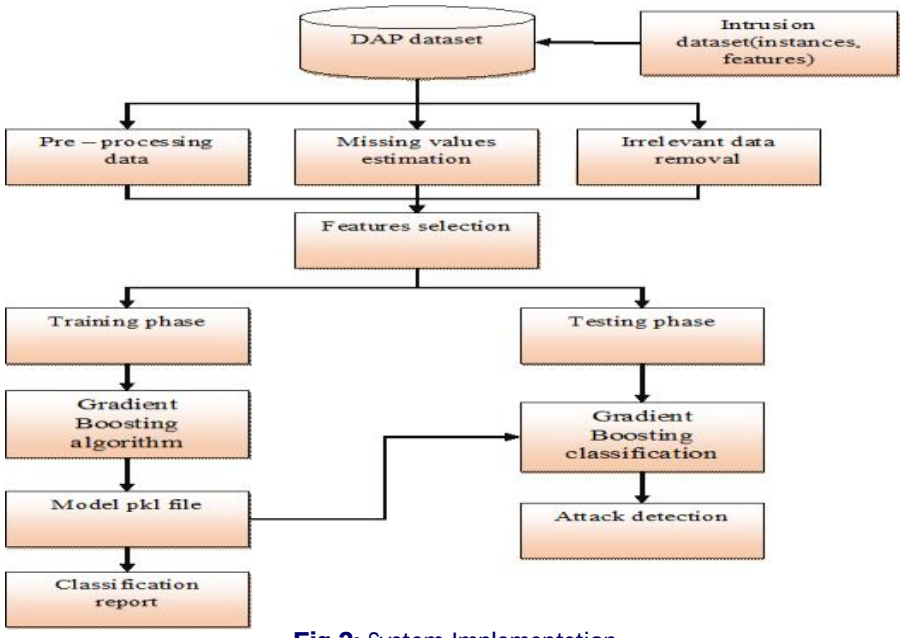

Fig.2: System Implementation

The training phase employs the Gradient Boosting algorithm, an ensemble learning method that sequentially builds decision trees to correct errors from previous iterations. In the testing phase, the saved model performs real-time classification on unseen network traffic, effectively distinguishing between benign activities and various APT attack stages proactive threat mitigation as shown in Fig. 2.

The Use Case Diagram illustrates the interaction between users and the APT Detection System. The Network Administrator uploads network flow data into the system, which is automatically processed and analyzed using machine learning algorithms. The Security Analyst reviews the detection results and generates security reports. This diagram helps in clearly understanding the responsibilities of each actor and the functional capabilities of the system as shown in Fig 3.The performance of the proposed APT detection system was evaluated using a confusion matrix, as shown in Fig. 4. The matrix presents the classification results across five categories: benign traffic, data exfiltration, foothold, lateral movement, and reconnaissance. The diagonal elements represent correctly classified instances, while off-diagonal values indicate misclassifications.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

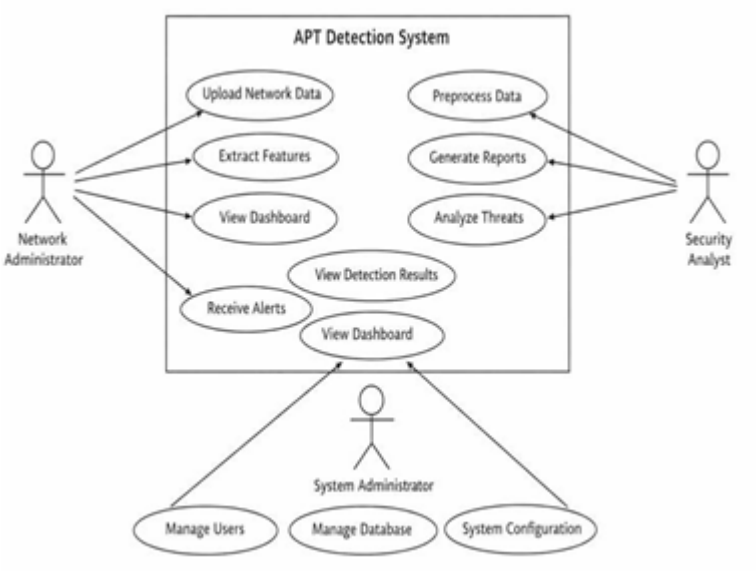

Fig.3 Usecase Diagram

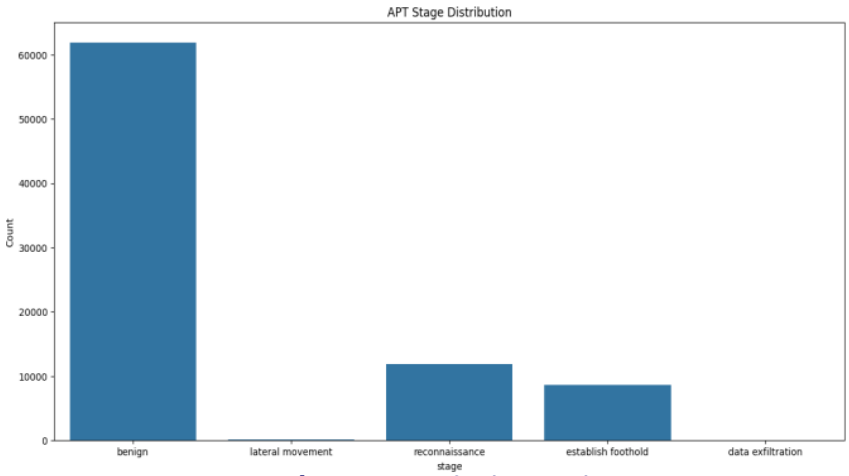

Fig.4 APT Confession Matrix

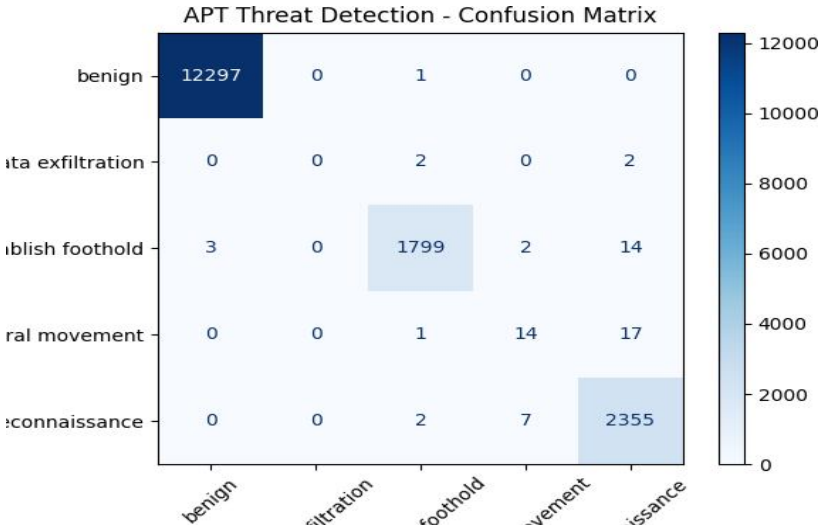

Fig.5 APT Stages Graph

The distribution of detected APT stages across the analyzed network traffic is shown in Fig. 5, which illustrates the frequency counts for five distinct categories: benign traffic, lateral movement, reconnaissance, establish foothold, and data exfiltration. As shown, benign traffic dominates the dataset with the highest frequency, reflecting the realistic class imbalance inherent in enterprise network environments where normal operations vastly outnumber malicious activities scanning and information gathering as initial phases of APT campaigns.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

Key Considerations:

Data Quality and Availability: The collected network flow data must be accurate, complete, and properly labeled to ensure reliable threat detection. Poor quality or incomplete data may reduce the effectiveness of the detection system.

Feature Selection: Selecting the most relevant network flow features is critical for improving detection accuracy. Features such as packet size, session duration, protocol type, and traffic frequency should be carefully chosen to help the machine learning model distinguish between normal and malicious network behavior.

Model Accuracy and Performance: The machine learning algorithms used in the system must be optimized to achieve high detection accuracy while minimizing false positives and false negatives. Continuous model evaluation and tuning are required to ensure reliable identification of Advanced Persistent Threat activities.

Real-Time Detection Capability: APT attacks often occur over long periods and involve stealthy communication patterns. Therefore, the detection system must support real-time or near real- time analysis of network traffic to quickly identify suspicious activities and generate alerts.

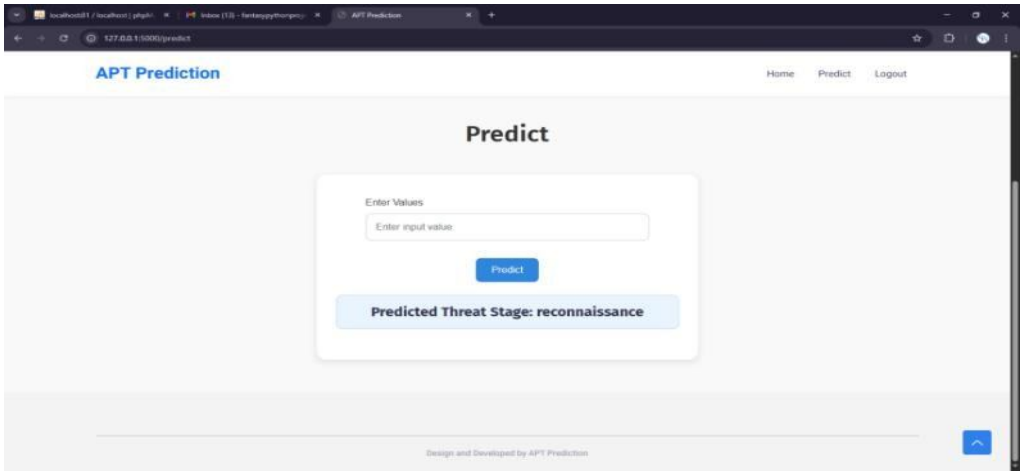

Fig 6: Prediction Page

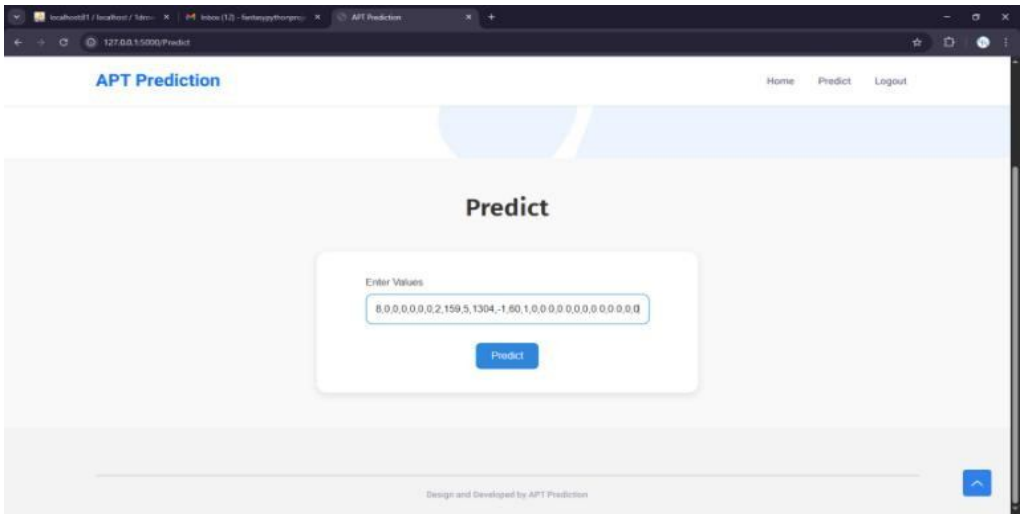

Fig 7: Threat Detection

CONCLUSION

The proposed system presents an effective approach for detecting Advanced Persistent Threat (APT) attacks using machine learning techniques and network flow analysis. By analyzing network traffic patterns and extracting important flow features, the system is able to identify abnormal activities that may indicate potential cyber threats. Machine learning algorithms improve the accuracy and efficiency of threat detection by automatically learning patterns from network data. The integration of network monitoring tools, data processing techniques, and visualization dashboards enables administrators to monitor network behavior and detect suspicious activities in a timely manner. The experimental results demonstrate that the proposed system can successfully classify network traffic as normal or malicious and provide early detection of potential security threats. Overall, the system contributes to improving network security by providing an intelligent and scalable solution for identifying Advanced Persistent Threat attacks.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.18

REFERENCES

1. P.Toupas,D.Chamou,K.M.Giannountakis,A.Drosou,and D.Tzovaras, "An intrusion detection system for multi-class classification based on deep neural networks," in 201918th IEEE International Conference on Machine Learning and Applications (ICMLA), 2019, pp.1-6.
2. P.C.Tikekar, S.S.Sherekar, and V.M.Thakre, "Features representation of botnet detection using machine learning approaches," in2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC),2020,pp.1-5.
3. D.Ucci,F.Sobrero,F.Bisio,andM.Zorzino,"Near-real-time anomaly detection in encrypted traffic using machine learning techniques," in 2022 IEEE International Conference on Cyber Security and Resilience (CSR), 2022, pp. 1-8.
4. B.YangandD.Liu,"Research on network traffic identification based on machine learning and deep packet inspection," in 2019 IEEE 3rd Information Technology,Networking, Electronic and Automation Control Conference (ITNEC),2019, pp. 1-5.
5. H.Dong,A.Munir,H.Tout,andY.Ganjali, "Next-generation data center network enabled by machine learning: Review, challenges, and opportunities, "IEEE Access, vol. 9, pp. 1-17, 2021.
6. R.Kaur, J.K.Sandhu, and L.Sapra, "Machine learning technique for wireless sensor networks," in 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 1-4.
7. D.Szostak, K.Walkowiak, and A.Wlodarczyk, "Short-term traffic forecasting in optical network using linear discriminant analysis machine learning classifier," in 2021 International Conference on Optical Network Design and Modeling (ONDM), 2021, pp. 1-4.
8. Y.Zhao,Y. Li,X. Zhang,G. Geng,W.Zhang,and Y.Sun,"A survey of networking applications applying the software defined networking concept based on machine learning,"IEEE Access,vol. 7, pp. 1-21, 2019.
9. P.Philipp,R.X.M.Georgi,J.Beyerer,andS.Robert,"Analysis of control flow graphs using graph convolutional neural networks,"in2019 6th International Conference on Soft Computing & Machine Intelligence (ISCMI), 2019, pp. 1-5.
10. Q.Wang, H.Yan, and Z.Han, "Explainable APT attribution for malware using NLP techniques," in 2021 IEEE 21st International Conference on Software Quality, Reliability and Security (QRS), 2021, pp. 1-11.
11. H.Bian,T.Bai,M.A.Salahuddin,N.Limam,A.A.Daya,R.Boutaba,"Host in danger Detecting network intrusions from authentication logs,"in 2019,15 International Conference on Network and Service Management(CNSM), 2019,pp.1-6.
12. O.McCusker, S.Brunza, and D.Dasgupta, "Deriving behavior primitives from aggregate network features using supportvectormachines,"in 2013 5thInternationalConference on Cyber Conflict (CYCON 2013), 2013, pp. 1-6.
13. Y.Xiuzhang,P.Guojun,L.Side,Z.Dongni,L.Chenguang, and L.Xinyi, "A survey on intelligent detection for APT attacks,"ChinaCommunications,vol.22,no.11,pp.103-131,Nov. 2025.
14. R.V.Umasevi and T.R.Nisha Dayana, "A hybrid technique for detecting cyber threats through network traffic analysis," in2025 5th International Conference on Expert Clouds and Applications (ICOECA), 2025, pp. 1-6.
15. S.Balaba, Y.Chernyshov, A.Skorohodov, and D.Komarov,"Graph-based anomaly detection in industrial control systems," in*2025 IEEE Ural-Siberian Conference on Biomedical Engineering, Radio electronics and Information Technology (USBEREIT)*, 2025, pp. 1-4.
16. S.Aruna, G.L.Prakash, and K.B.Surekha, "Advanced persistent threat detection system using network traffic analysis,"in2025 International Conference on Electronics and Computing, Communication and Networking Automation Technologies (ICEC2NT), 2025, pp. 1-6.
17. S.Muthumanikandan, G.K.Hegde, P.Swetha, and P.B.Honnavalli, "Flow-based behavioral analysis with machine learning for SSH lateral movement detection,"in2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA),2025, pp. 1-5.
18. K.Rathor, V.Keerthika, K.Sunanda, K.Renuga, A.Shobana,andM.Anusuya,"Enhancing network security against APTs through SVM-based network traffic analysis: Identifying anomalies in communication flows," in 2024 International Conference on Computing and Data Science(ICCDS),2024,pp. 1-5.
19. N.H.A.Mutalib,A.Q.M.Sabri,A.W.A.Wahab,E.R.M.F.Abdullah, and N.AlDahoul, "An explainable recursive feature elimination to detect advanced persistent threats using random forest classifier,"in 2025 3rd International Conference on Cyber Resilience (ICCR), 2025, pp. 1-6.
20. Q.Hu,Y.Wang, Z.Su, T.H.Luan, R.Li, and Z.Jiang, "Rethinking online smart contract diagnosis in blockchains: A diffusion perspective, "IEEE Transactions on Networking, vol. 34, pp. 230-245, Sep. 2025.