WatchDogX – AI Powered Cyber Threat Detection and Respnse System

Submitted by: Mridul Goyal

Mentor: Dr. Abhishek Jain

Institution: Bml Munjal University

Made with GAMMA

Problem Statement

Traditional intrusion detection system (IDS) solutions are signature-based and outdated, unable to detect novel or slightly modified attacks. Manual analysis of security logs is time-consuming and error-prone. Additionally, most open-source IDS tools lack real-time alerting capabilities, leaving networks vulnerable to emerging threats.

Made with GAMMA

Our Solution Objectives

1

AI-Driven Real-Time Threat Detection

Develop an intelligent system capable of identifying cyber threats instantly using advanced AI algorithms.

2

Flexible Data Ingestion

Support both live packet sniffing for dynamic monitoring and manual CSV uploads for forensic analysis.

3

Multi-Channel Alerting

Integrate real-time alerts via SMS and Discord to ensure rapid response and immediate stakeholder notification.

4

Actionable Visualizations

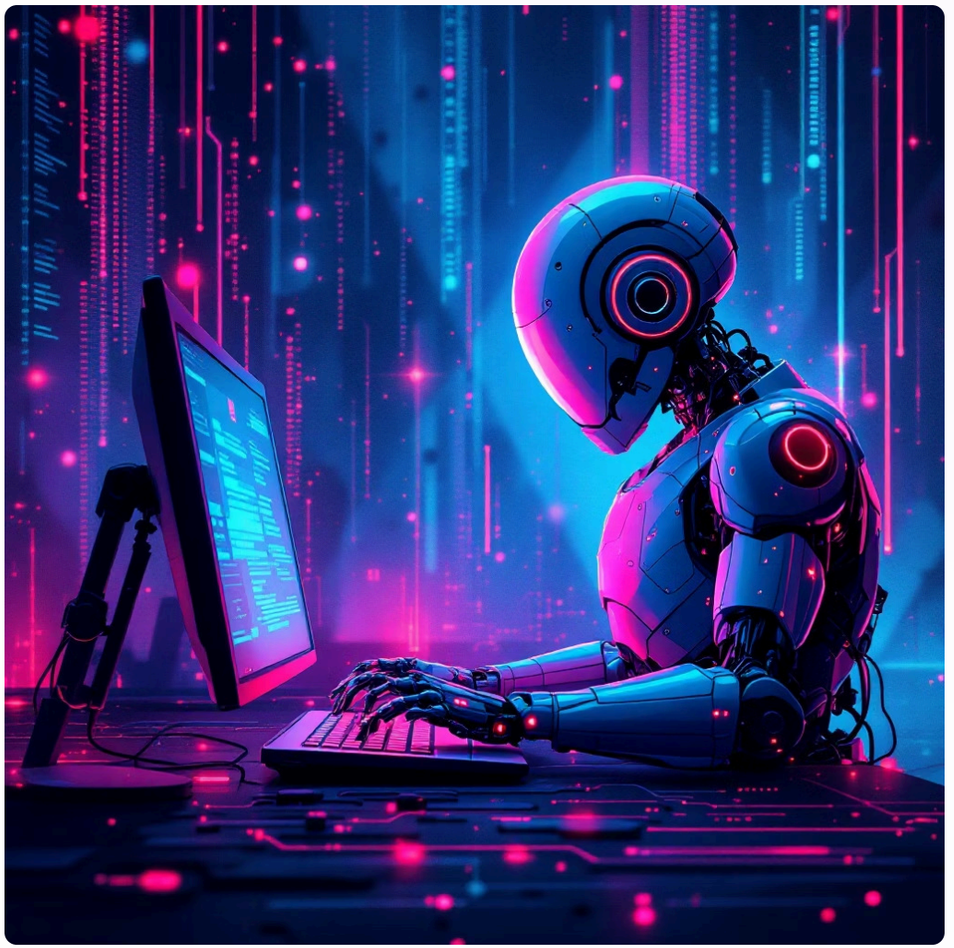
Provide intuitive dashboards and geolocation maps for clear visualization of incidents and threat patterns.

Made with GAMMA

Motivation: The Evolving Threat Landscape

Rise of Automated Attacks

The proliferation of sophisticated, automated cyber attacks necessitates a defense system that can adapt and respond dynamically, moving beyond static, signature-based approaches.



Limitations of Traditional Systems

Outdated signature-based Intrusion Detection Systems (IDS) are increasingly ineffective against polymorphic and zero-day threats, highlighting the urgent need for proactive, AI-driven defenses.

Demand for Accessible Solutions

There is a significant need for affordable and lightweight real-time threat detection tools, making advanced cybersecurity accessible to general users and small to medium-sized enterprises.

Made with GAMMA

Literature Review: Foundations and Innovations

CICIDS2017 Benchmark: The CICIDS2017 dataset, as established by Sharafaldin et al., is a cornerstone for modern IDS research due to its realistic network traffic simulations and diverse attack scenarios.

ML Superiority: Machine Learning (ML) models consistently outperform traditional rule-based and signature-based systems in identifying novel and evolving cyber threats, demonstrating enhanced adaptability and accuracy.

Optimization Techniques: Techniques like undersampling and meticulous feature engineering significantly improve model performance by addressing class imbalance and extracting the most relevant data characteristics.

Current Gap: Existing cybersecurity tools often lack integrated real-time detection with comprehensive, multi-channel alerting capabilities, a critical gap that WatchDogX aims to fill.

Made with GAMMA

Dataset: CICIDS2017

- **Comprehensive Size:** Approximately 12 GB of raw network traffic, offering a robust foundation for training and validation.
- **Diverse Attacks:** Includes 14 distinct attack types alongside legitimate (BENIGN) traffic, enabling comprehensive model training.
- **Structured Format:** Raw PCAP files were systematically filtered, cleaned, and converted into CSVs, then categorized by specific attack types for targeted analysis.
- **Class Balancing:** The dataset was balanced using RandomUnderSampler to prevent model bias towards majority classes, ensuring fair representation of all attack types.

Made with GAMMA

Data Preprocessing: Refining the Raw Data

Raw Traffic Merging & Cleaning

Initial phase involved consolidating and sanitizing diverse network traffic logs to ensure data integrity and consistency.

Attack-wise File Segregation

Traffic was then meticulously separated into attack-specific CSV files (e.g., "Attack_vs_BENIGN.csv") for focused model training.

Feature Selection

Critical features were identified using correlation analysis, selecting the top 3-4 most impactful attributes per attack type to optimize model efficiency.

Key Common Features

Common features consistently found across various attack types, such as Flow Duration, Packet Length Std, and Bwd Packet Length Mean, were prioritized for their predictive power.

Made with GAMMA

Machine Learning Models: Selection and Implementation

Our rigorous model selection process evaluated several algorithms to identify the most effective for real-time threat detection. The **Multi-Layer Perceptron (MLP)** emerged as the final, most robust model due to its high accuracy and adaptability to diverse network traffic patterns.

- **Multi-Layer Perceptron (MLP):** Selected as the primary model for its superior performance and generalization capabilities.
- **Naive Bayes:** Evaluated for its probabilistic approach and computational efficiency.
- **Quadratic Discriminant Analysis (QDA):** Considered for its ability to model complex decision boundaries.

Post-training, each model and its corresponding scaler are meticulously serialized into **.pkl** files using **joblib**, facilitating seamless integration and deployment within the WatchDogX system.

Made with GAMMA

Model Performance: Achieving High Accuracy

The Multi-Layer Perceptron (MLP) model consistently demonstrated exceptional performance across various attack types, validating its suitability for the WatchDogX system.

96-99%

0.95-0.98

MLP Accuracy

Consistently achieved high accuracy rates, validating its effectiveness in classifying malicious and benign network traffic.

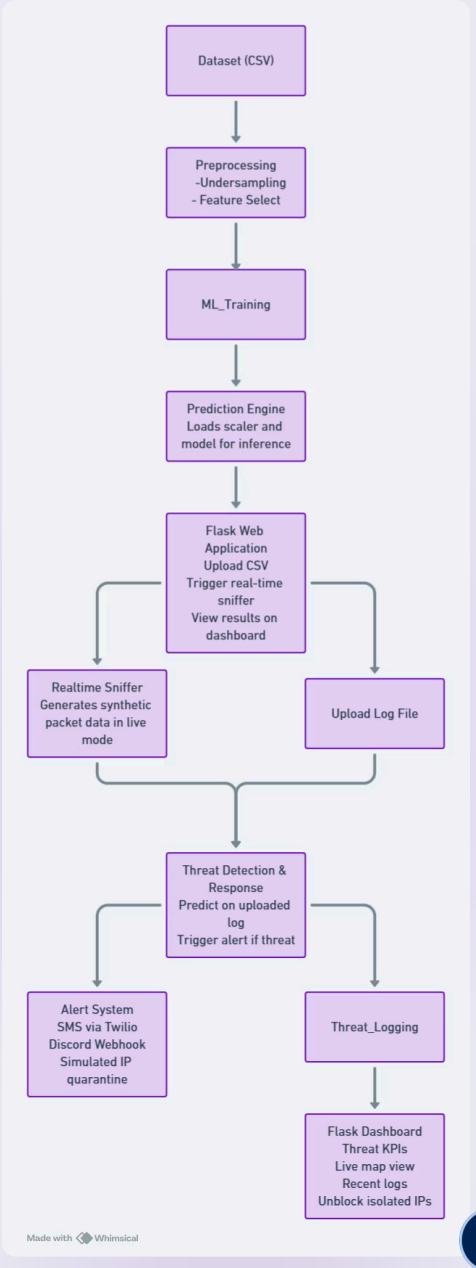
High F1-Scores

Particularly strong performance in detecting DDoS, Botnet, and DoS attacks, demonstrating high precision and recall.

	Attack Type	Naive Bayes Accuracy	QDA Accuracy	MLP Accuracy
0	Bot	0.638021	0.832589	0.998870
1	DDoS	0.984610	0.980481	0.992264
2	DoS GoldenEye	0.984251	0.984238	0.996942
3	DoS Hulk	0.977031	0.976215	0.989537
4	DoS Slowhttptest	0.972274	0.974597	0.996782
5	DoS slowloris	0.981212	0.978324	0.997528
6	FTP-Patator	0.998758	0.998777	0.995455
7	PortScan	0.916335	0.916335	0.998157
8	SSH-Patator	0.998276	0.998788	0.998808

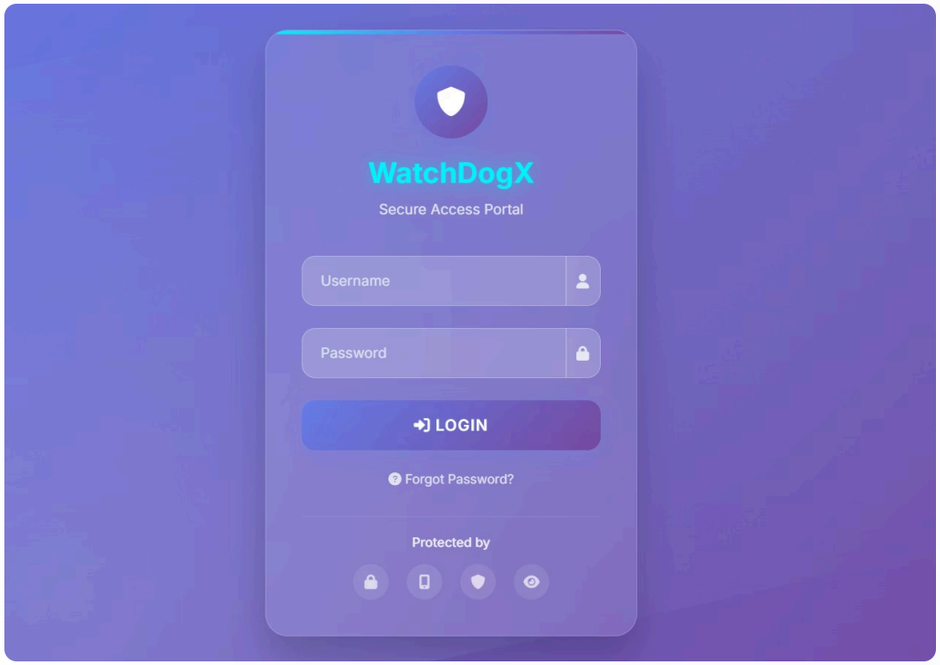
Made with GAMMA

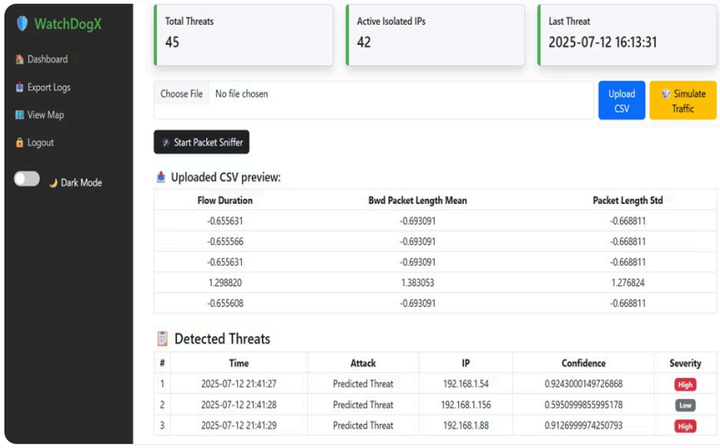
Architecture Diagram: WatchDogX System Flow



Made with GAMMA

Flask Web Dashboard: User Interface





- **Secure Authentication:** Robust user authentication mechanisms ensure secure access to the dashboard and its sensitive functionalities.
- **CSV Prediction Upload:** Users can easily upload network traffic CSV files for immediate threat prediction and analysis.
- **Live Packet Sniffer:** Integrated functionality to initiate real-time packet sniffing for on-the-fly threat detection.
- **Comprehensive Threat Insights:** Displays detailed threat logs, IP isolation capabilities, and an interactive threat map for quick operational understanding.
- **Dark Mode UI:** Features a sleek dark mode user interface, built with Bootstrap, for enhanced user experience and reduced eye strain during prolonged monitoring.

Made with GAMMA

Integrated Alerting System

SMS Notifications

Integrated with **Twilio** for instant SMS alerts, ensuring critical security notifications reach personnel even on the go.

Discord Webhooks

Leverages **Discord Webhooks** for real-time notifications, facilitating quick team collaboration and response in dedicated security channels.

Threshold-Based Triggers

Alerts are dynamically triggered based on the AI model's confidence threshold, ensuring timely and relevant threat awareness.

Automated Actions

Configurable actions include IP quarantine, comprehensive event logging, and immediate user notifications for swift incident mitigation.

Made with GAMMA

Intuitive User Workflow

1

Login

Secure user authentication to access the platform.

2

Data Ingestion

Upload CSV files for retrospective analysis or activate the live packet sniffer for real-time monitoring.

3

View Predictions

Access real-time threat predictions via interactive dashboards.

4

Alerts & Logs

Receive instant alerts, with all events securely logged and geo-tagged.

5

Management

Optionally unblock isolated IPs, download comprehensive CSV reports, and view dynamic threat maps.

Made with GAMMA

Rigorous Testing & Validation

Unit Testing

Comprehensive unit tests cover core functionalities, including model prediction accuracy and robust CSV parsing.

Integration Testing

Thorough integration tests validate seamless interactions between alerts, logging mechanisms, and Flask API endpoints.

Edge Case Analysis

Extensive testing addresses challenging scenarios, such as missing data values and unbalanced input streams, ensuring system resilience.

Hybrid Validation

Both manual and synthetic testing methodologies are employed to rigorously validate the system's performance and reliability.

Made with GAMMA

Key Features & Future Outlook

Key Features

- Real-time AI-powered threat detection
- Dual-mode operation: CSV upload and live sniffer
- Multi-channel alerting (SMS, Discord)
- Geo-logging and interactive threat dashboard
- Easy CSV export and visualization tools

Future Work

- Direct integration with enterprise firewall systems
- Implement SHAP for explainable AI insights
- Develop role-based access control (RBAC)
- Transition to a cloud-hosted, continuous service

Made with GAMMA

Thank You

Made with GAMMA