Applying AI/ML in DDoS Traffic Analysis and Prediction

From Reactive to Proactive Defense

VNIX-NOG CONFERENCE

October 17th,2025

LONG VÂN SYSTEM SOLUTION
PHẠM TRƯỜNG VŨ — Network Engineer

LONG VAN System Solution

The Evolving Threat Landscape

Sophistication of Modern DDoS Attacks

- Started with simple volume floods (SYN/UDP).
- Evolved to low-and-slow, HTTP floods.
- Now includes SSL DDoS and application-layer attacks.

LAYER 3 | LAYER 7 | MULTI-VECTOR

Inadequacies of Traditional Methods

- Signature-based systems miss zero-day threats.
- Static thresholds generate false positives/negatives.

Flash Crowd vs. Attack Dilemma

Legacy systems struggle to distinguish legitimate traffic surges from malicious attacks, leading to either blocking legitimate users or allowing breaches.

OUTDATED | REACTIVE | INFLEXIBLE

The Evolving Threat Landscape

"To effectively combat these threats, adaptive, learning-based systems are no longer optional – they are essential."

Journal of Artificial Intelligence & Cloud Computing (2024)

Adaptive vs. Static Thresholds: Responding to Traffic Anomalies

From Traffic to Data: The Network Engineer's Goldmine

Raw Network Flow: The Foundation

Access comprehensive network visibility from diverse data sources:

- NetFlow/IPFIX
- sFlow
- SNMP
- Syslog
- BGP Logs

AI/ML: Unlocking Actionable Insights

Advanced AI/ML models transform raw network flow data into proactive security and performance insights.

Cleaning & Transformation Imperative

Raw data reveals baselines and anomalies, demanding rigorous cleaning and transformation for actionable intelligence.

Flow Metadata: Lightweight & Scalable

Embrace flow metadata for efficient, large-scale network monitoring. Gain critical insights without the overhead of full packet analysis.

Reference:

DataBank: How AI is Transforming the Battle Against DDoS Attacks

AI/ML Demystified for Network Security

Machine Learning (ML) finds **patterns** in data, **without explicit rules**.

Primary Function

Anomaly Detection: Identifies unusual behaviors on time-series & feature vectors.

- Time-series Analysis
- Feature Vector Analysis

Learning Approaches (Brief)

- Supervised Learning
- Unsupervised Learning

*Details on the Next Slide

Network Data Inputs

Packets/sec | Bytes/sec | Unique Sources | TCP Flags | Geolocation

ML Processing Engine

Algorithms analyze data streams to detect deviations from the norm.

Actionable Outputs

Anomaly Score | Classification Label | Predicted Trend

Supervised vs Unsupervised: Choosing the Right Tool

Supervised Learning: Guided Accuracy

- Requires Labeled Data
- Accurate with Large Datasets
- Predicts Known Outcomes

> Learns from pre-defined input-output pairs.
>
> Excellent for detecting known attack types.

Examples: **Random Forest, XGBoost**

Known Threats | Data Driven

Unsupervised Learning: Pattern Discovery

- Learns "Normal" Patterns
- Detects Zero-Day Threats
- No Labeled Data Needed

> Identifies hidden structures in unlabeled data.
>
> Crucial for evolving threat landscapes.

Examples: **IsolationForest, Autoencoder**

Emerging Threats | Pattern Recognition

Hybrid Approach: Strength in Combination

Combine supervised (knowns) and unsupervised (unknowns) models with a human-in-the-loop to reduce false positives and enhance detection accuracy.

The ML Pipeline for DDoS Detection

Data Collection

- NetFlow / sFlow export
- Centralized data collector

Traffic Telemetry · Network Flow

Feature Engineering

- Source count, packet rate
- Flag ratios, entropy
- Geo-distribution

Critical for Accuracy

Model Training

- Historical data analysis
- Cross-validation & Tuning

Learning Phase

Inference

- Real-time / Near-real-time scoring
- Anomaly detection

Live Detection · Rapid Analysis

Action

- Alerts / Automated mitigation
- ACLs, Firewall APIs, Scrubbers

Automated Response

Feedback Loop

- Retrain with new labeled data
- Continuous adaptation

Refinement · Evolving Threats

"Harnessing AI for Network Security and DDoS Attack Detection offers significant advancements in combating evolving threats."

Source: Harnessing AI for Network Security and DDoS Attack Detection - J Arti Inte & Cloud Comp, 2024

Key Technique: Traffic Baseline (Time Series)

Core Objective: Identify Baseline Behaviors

- Learn traffic patterns by hour, day, and week.
- Establish a normal traffic baseline for analysis.

Network Telemetry

Predictive Analytics

Leveraging AI for Anomaly Detection

Models Employed:

ALGORITHM ARIMA

Prophet

Application:

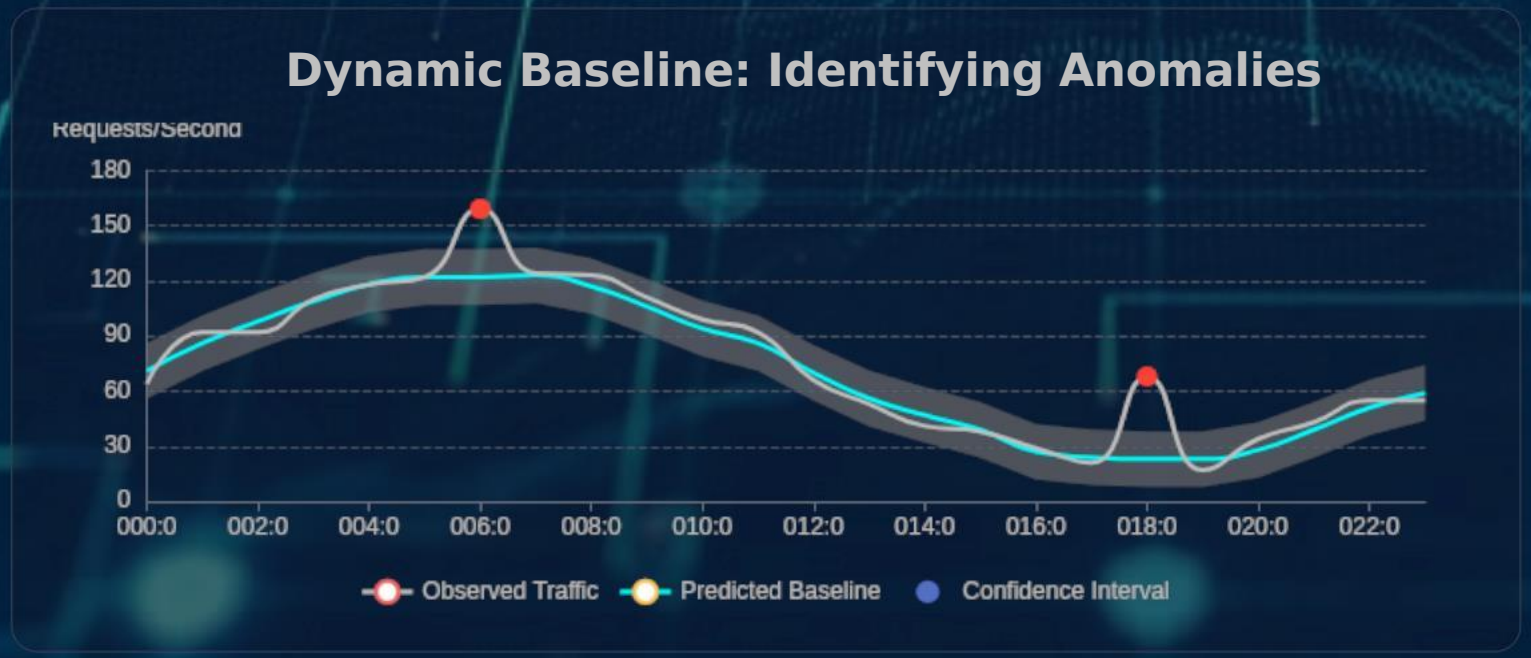
Detect spikes vs. predicted baseline

Utilize adaptive thresholds

AI Security

Proactive Defense

Key Technique: Traffic Baseline (Time Series)



Key Recommendation

"For superior detection, leverage **confidence intervals** instead of rigid, fixed thresholds."

Adaptive Security

Threshold Optimization

Source: How AI is Transforming the Battle Against DDoS Attacks - DataBank
Source: Harnessing AI for Network Security and DDoS Attack Detection

Key Technique: Supervised Classification

Core Objective

- Accurately distinguish Normal vs. DDoS traffic.
- Enable granular identification of attack subtypes.
- **Impact: Strengthens cyber defense strategies.**

Network Security | Threat Intelligence

RandomForest

- Ensemble method using multiple decision trees.
- Robust for high-dimensional tabular features.
- Reduces overfitting for better generalization.

Ensemble Learning

XGBoost

- Optimized Gradient Boosting framework.
- Exceptional performance on structured datasets.
- Supports parallel computation for efficiency.

Gradient Boosting

Essential Prerequisites

- **Labeled Dataset:** Foundation for training (e.g., CIC-DDoS2019).
- **Cross-Validation:** Critical to prevent overfitting and ensure reliability.

Sources: J Arti Inte & Cloud Comp (2024), DataBank (2025)

Key Technique: **Unsupervised Anomaly Detection**

Isolation Forest

Identifies anomalies by isolating data points.

Effective for high-dimensional datasets.

Leverages unique tree-based partitioning.

- Novel Threat Detection

Autoencoders

Learns baseline of normal network traffic.

Flags unusual patterns via high reconstruction error.

Adaptive to evolving network behaviors.

- Abnormal Traffic Signals

First Defense Layer: Zero-Day Attacks

Crucial for detecting unknown and emerging threats.

Acts as a proactive barrier against novel attacks.

Minimizes dwell time for advanced adversaries.

Threshold Tuning & Monitoring

Fine-tuning detection thresholds is critical.

Continuous monitoring reduces false positives.

Key Technique: Clustering for Attack Analysis

Core Concept

Leverage **K-Means** and **DBSCAN** algorithms to group source IPs by their unique network behaviors.

- K-Means
- DBSCAN
- IP Behavioral Analysis

Primary Objectives

- Identify botnets and coordinated attacks
- Detect abnormal request patterns
- Analyze TTLs and window sizes for anomalies

Visualizing distinct clusters of network activity.

- Grouping & Classification
- Bulk Mitigation
- Digital Forensics

Anomalous clusters, such as one generating **70% of traffic**, demand immediate investigation for effective threat response.

Bringing It All Together: Multi-Layered AI Defense

Real-time Unsupervised Detection

Immediate anomaly identification through advanced unsupervised AI.

Speed & Autonomy

Near Real-time Classification

Contextual validation with supervised models to minimize false positives (FP).

Accuracy & Reduction

Analysis & Forensics

Deep dive investigation, threat clustering, and attack attribution.

Insight & Attribution

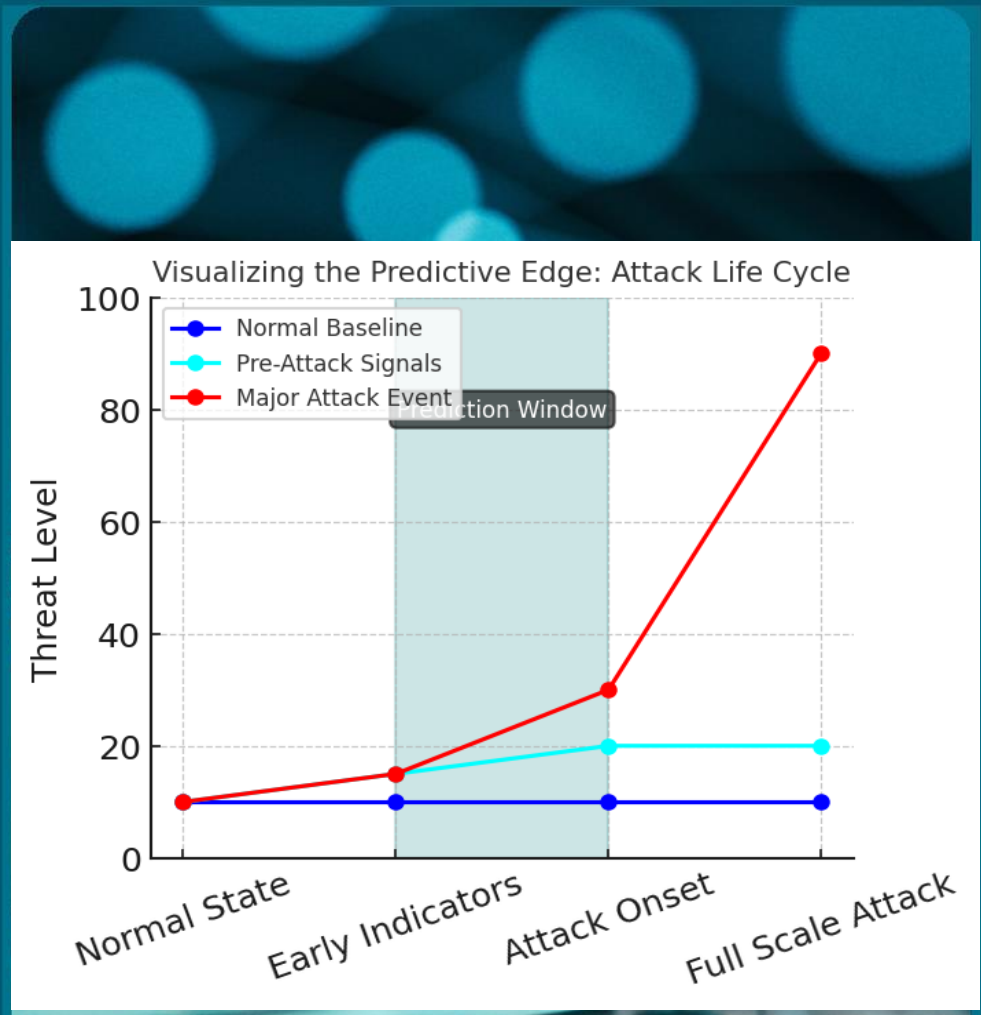
Our defense is a **strategic fusion** of advanced AI and human expertise, delivering true **defense-in-depth**.

Orchestration Core

Seamlessly integrating SIEM/SOC with intelligent playbooks and automation APIs for synchronized, efficient incident response.

Automated Workflows | Adaptive Playbooks

From Detection to Prediction



The Predictive Leap

Transitioning from reactive defense to proactive cyber threat anticipation.

Identify subtle, pre-attack behaviors to neutralize threats before impact.

Key Indicators

- SYN Scan Bursts
- Probe Activity Spikes
- Cross-Source Anomalies

Detecting unusual network patterns across diverse data streams.

Methodologies

- Historic Context Analysis
- Multi-Signal Correlation
- Advanced Temporal Models

Leveraging past attack patterns and real-time data for foresight.

Current Landscape & Application

A promising field, actively undergoing extensive research.

Initial deployment focuses on generating pre-emptive **advisory alerts**.

> ⚠ *Example: A significant increase in SYN scans often precedes a large-scale DDoS flood attack.*

Architecture in Practice

Deployment Strategy

Flexible & controlled rollout.

- On-premise
- Cloud-based

- Offline pre-automation testing recommended
- Phased automation rollout

Key Integration Points

Ensuring seamless system interoperability.

- Standardized Alert Format (JSON)
- Robust API Access for Actions
- Managed API Rate Limits

What You Can Do Next Monday

1. Enable network Flow data

Activate NetFlow/IPFIX on core routers to begin rich traffic data export to your collector.

- NetFlow/IPFIX
- Core Router

2. Visualize & Build Baseline

Utilize ELK Stack or Grafana to visualize traffic patterns and establish normal behavior baselines.

- ELK/Grafana
- Network Baseline

3. Pilot Protocol Analysis

Select a single protocol, such as DNS, for a focused pilot to build and validate an initial anomaly detection model.

- Pilot Protocol (DNS)
- ML Model Validation

4. Develop Offline POC

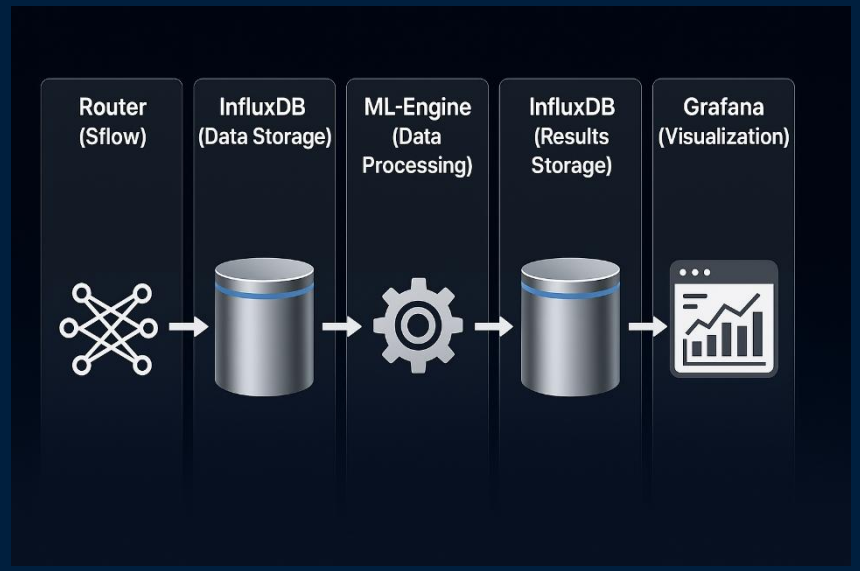
Create a Jupyter Notebook proof-of-concept for the model in an offline environment before real-time deployment.

- Jupyter Notebook
- Offline POC

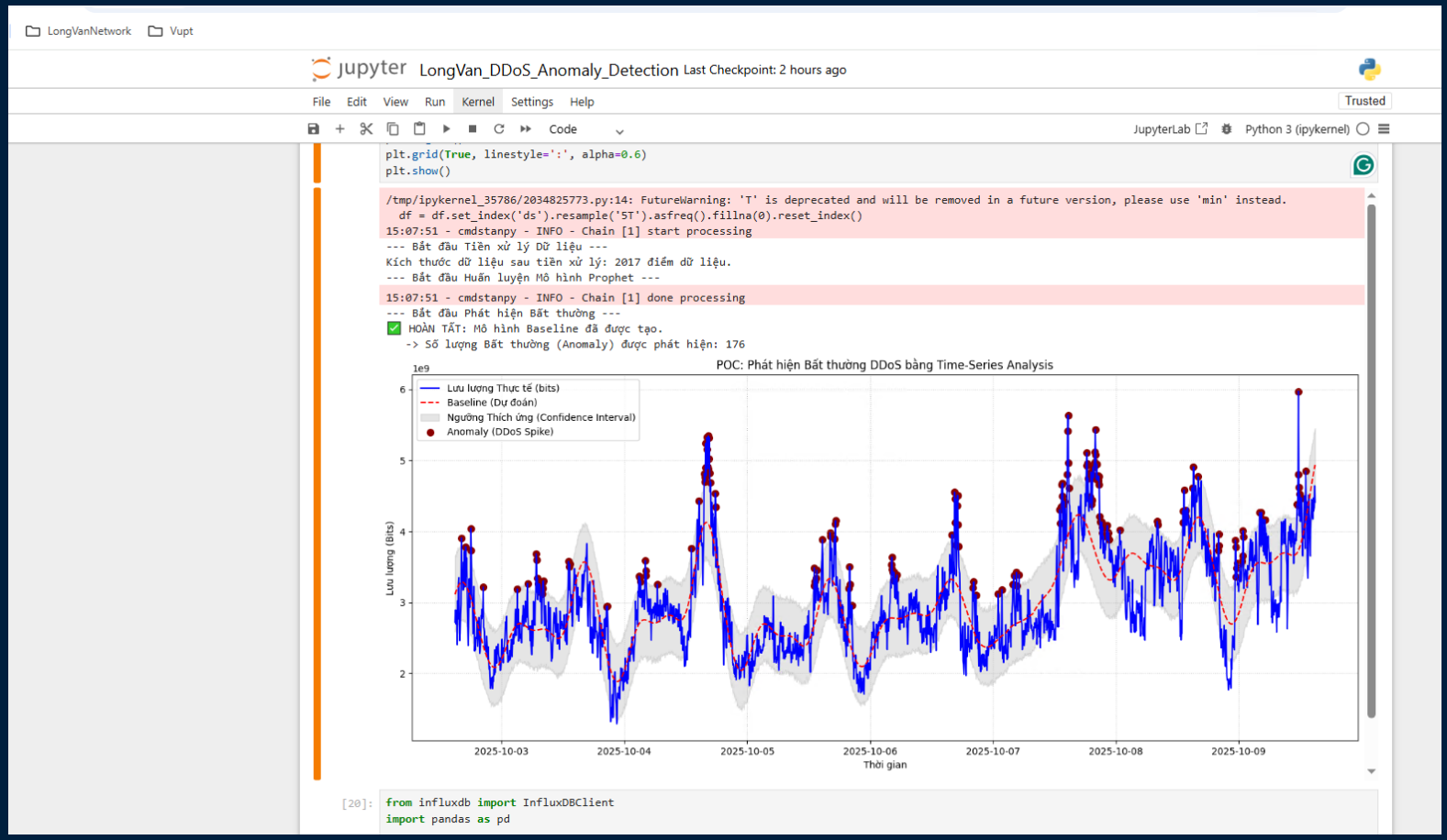
Source: CEO Monthly, MDPI, HAL Science

Long Van Case Study: Unsupervised Learning for Adaptive DDoS Anomaly Detection

The ML Pipeline



Jupyter Note book Environment



Long Van Case Study: Unsupervised Learning for Adaptive DDoS Anomaly Detection

Grafana Dashboard: Adaptive DDoS Threshold Monitoring

Challenges & Considerations

Data Quality & Quantity

- Need for extensive data volume.
- Data must be clean and deduplicated.
- Requires enrichment (Geo-location, ASN).

"Quality of AI outcomes hinges on data."

Clean Data · De-duplication · Geo/ASN Enrichment

False Positives & Negatives

- Risk of misidentifying attacks or missing real threats.
- Requires human-in-the-loop validation.
- Crucial for continuous feedback loops.

Anomaly Tuning · Expert Review · Human Oversight · Feedback Mechanism

Interdisciplinary Expertise

- Demands synergy between network engineers.
- And data science specialists.
- To interpret and act on insights.

Cross-Functional Teams · Skill Development · Network Engineering · Data Science

Resource & Privacy Constraints

- Significant compute and storage costs.
- Strict data compliance (GDPR, ND).
- Balancing utility with data protection.

Cost Optimization · Secure Storage · Compliance · Down-sampling

Conclusion & Q&A

The Evolving DDoS Threat

DDoS attacks are increasingly sophisticated — static defenses aren’t enough.

Advanced Attacks **Insufficient Defenses**

AI/ML: Essential for Modern Defense

- Detection: Pinpoint anomalies with precision.
- Classification: Accurately categorize attack types.
- Prediction: Anticipate and pre-empt future threats.

Proactive Intelligence

Source: DataBank

Strategic AI/ML Integration

Phase 1: Start small: Pilot programs for initial efficacy validation.

Phase 2: Validate: Measure effectiveness and refine algorithms.

Phase 3: Scale: Expand successful solutions across the network.

Phase 4: Integrate: Seamlessly merge with existing mitigation strategies.

Pilot **Validate** **Scale** **Integrate**

Your Next Step: Immediate Action

Let’s start by enabling NetFlow this week.

NetFlow **Actionable Insight**

Open Discussion & Contact

We welcome your questions and are ready to discuss your specific cybersecurity needs.

For further inquiries, reach out at x@y.com

Q&A Session **Connect**