WATCH TOWER - ZEEK ML PIPELINE

🔥 Why This is a New Model Compared to Existing Ones?

1. **Combination of Signature-Based & Anomaly-Based Detection**
    - Most systems either use **signature-based** (e.g., Snort, Suricata) or **anomaly-based** (e.g., UEBA models).
    - Watch Tower – Zeek ML pipeline **combines both** using **Zeek logs + Random Forest (for known attacks) + Isolation Forest (for unknown threats)**.
    - This hybrid model increases accuracy and adaptability.
2. **Automated Weekly Behaviour Analysis**
    - Many SIEM solutions detect threats **in real-time** but do not **analyze long-term behaviour**.
    - Watch Tower ML model **compares Week 1 to Week 2, then Week 3 to Week 2, and so on**, identifying evolving attack patterns.
3. **Remote & Scalable Monitoring**
    - Unlike traditional IDS/IPS (which monitor local traffic), your system **collects logs from remote machines** and analyses them centrally.
    - This is useful for large-scale **enterprise networks & cloud-based environments**.
4. **ELK Stack Integration for Real-Time Threat Intelligence**
        - Real-time log storage
        - Threat visualization
        - Immediate alerting

5. **Supports Continuous Learning & Adaptation**
    - The **Isolation Forest model retrains** as new logs are ingested, allowing the system to evolve and detect **new types of threats over time**.
    - Can be further improved with **active learning (semi-supervised ML).**

How is This Better Than Existing Cybersecurity Models?

Feature	Traditional IDS/IPS (Snort, Suricata)	UEBA & Anomaly Detection	Watch Tower Zeek ML Pipeline
Detects Known Threats	Yes (Signatures)	No	Yes (Random Forest)
Detects Unknown Threats	No	Yes (Behavioural Analysis)	Yes (Isolation Forest)
Remote Log Collection	No	No	Yes
Automated Weekly Analysis	No	No	Yes
ELK Stack Integration	No	No	Yes

Is this a Completely New Model?

- Watch Tower is **not just a single ML model** but **a complete cybersecurity framework** that **combines Zeek + ML + ELK in a unique way**.
- While individual components exist (Zeek, Random Forest, Isolation Forest), their **integration and behaviour-driven approach make this a novel system**.
- If optimized further (e.g., deep learning, reinforcement learning), **this could be a groundbreaking cybersecurity solution.**

WATCH TOWER - ZEEK ML PIPELINE

Watch Tower updated pipeline is a more advanced and efficient approach compared to the initial ML model.

1. **Hybrid Threat Detection (Known + Unknown)**
    - Original model only used **Random Forest**, which relies on labeled attack data.
    - The updated version integrates **Isolation Forest**, which detects anomalies (zero-day threats).
    - This allows for both **pattern-based** and **behaviour-based** detection.
2. **Real-time & Scalable Monitoring**
    - **Zeek logs** provide deep network traffic analysis.
    - **Remote Log Collection** allows monitoring multiple machines, increasing scalability.
    - **ELK Stack Integration** enables centralized visualization & alerting.
3. **More Robust Analysis & Reporting**
    - Weekly reports generated with **detailed insights**.
    - Can be **correlated over time** to track evolving attack patterns.

🔥 **Why This is Unique & Powerful:**

- It **combines Zeek (Network Monitoring) + Machine Learning (Threat Detection) + ELK (Visualization & Alerting).**
- Instead of **just detecting known threats**, it **identifies unusual behaviour** that might indicate unknown or zero-day attacks.
- Can be **expanded to support more ML models (e.g., Autoencoders, Reinforcement Learning).**

WATCH TOWER - ZEEK ML PIPELINE

The ML models used in Watch Tower - Zeek ML Pipeline are:

1. **Random Forest Classifier (Supervised Learning)**
    - Used for detecting known attack patterns based on historical labelled data.
    - Trained on extracted features from Zeek logs.
    - Saves the trained model as rf_model.pkl.
2. **Isolation Forest (Unsupervised Learning)**
    - Used for detecting zero-day threats and anomalies in network behaviour.
    - Works by identifying outliers in network traffic.
    - Saves the trained model as iso_forest.pkl.

Downloads and Replacements:

1. **Dependencies are missing** - Ensure you have all required Python libraries installed. You can install them using:
2. **pip install –** pandas, numpy, joblib, matplotlib, fpdf, scikit-learn, requests, elasticsearch, paramiko
3. **Zeek logs are not in the expected path** – If Zeek logs are stored elsewhere, update the log_file path in the script.
4. **Elasticsearch is not running** – Ensure your Elasticsearch instance is correctly set up and running at http://localhost:9200.
5. **ML models don't exist yet** – If you're running the script for the first time, it will automatically train and save rf_model.pkl (Random Forest) and iso_forest.pkl (Isolation Forest). If you have pre-trained models, replace them accordingly.

6. **Remote Monitoring Setup** - If you're monitoring another PC remotely, ensure:
    - The remote machine has SSH enabled.
    - You have the correct SSH credentials.
    - The Zeek logs are accessible on the remote machine.

Report Format:

Weekly Security Report - Week [Week Number]

Threat Analysis:

[If threats were detected:] Threat ID: [Threat ID] | Source IP: [Source IP] | Destination IP: [Destination IP] | Protocol: [Protocol] | Service: [Service] | Connection State: [Connection State] | Confidence: [Confidence]% | Detection Method: [Detection Method] [Repeat the above line for each detected threat]

[If no threats were detected:] No threats detected this week.

User Behaviour Analysis:

[If user behaviour data is available:] User: [User IP] | Cluster: [Behaviour Cluster] | Total Duration: [Total Duration] hrs | Bytes Sent: [Bytes Sent] | Bytes Received: [Bytes Received]

[If anomalies are detected for the user:] Behavioural Anomalies Detected: [Anomaly Description] [Repeat the above for each user]

[If no user behavior data is available:] No significant user behavior anomalies detected.

Pattern Analysis and Insights:

[If threats were detected:] Threat ID [Threat ID] from Source IP [Source IP]: [Analysis/Insight based on threat details and confidence] [Repeat for each threat]

[If user behavior anomalies were detected:] User [User IP]: [Analysis/Insight based on user behavior and cluster] [Repeat for each user with anomalies]

[If no threats or user behavior anomalies were detected:] No significant user behavior detected.

Report Generated by ZREX AI Security System

DATASETS: CIC-IDS 2017, CIC-IDS 2018