CWL
CyberWarFare Labs

Anomaly Detection in Splunk using ML Model

About CW Labs :

CW Labs is a renowned UK based Ed-tech company specializing in cybersecurity cyber range labs. They provide on-demand educational services and recognize the need for continuous adaptation to evolving threats and client requirements. The company has these primary divisions:

1. Cyber Range Labs
2. Up-Skilling Platform
3. Infinity Learning Platform

CWL provides customized on-site training for organizations and accepts bulk orders for corporate upskilling. For inquiries or more details, feel free to reach out at **support@cyberwarfare.live.**

About Speaker

Rahul Chakraborty

(Security Engineer-Blue Team)

A Blue Team Security Engineer with hands-on experience in defensive security operations. His work primarily focuses on SIEM configuration, log onboarding, and detection engineering to enhance organizational visibility and security monitoring capabilities.

Agenda

- Problem Statement
- Why Traditional Security Fails
- Dataset Overview
- Why Encode Process Names
- Numeric Transformation Approach
- Smart Outlier Detection in Splunk
- Splunk MLTK Add-on Overview
- Demonstration Walkthrough

Problem Statement:

Security teams often face challenges:

1. High volume of process events
2. Unknown or living-off-the-land binaries
3. Signature-based detections missing rare behaviors
4. Difficulty spotting low-frequency malicious processes

Why does Traditional Security fail?

Traditional security controls rely heavily on signatures, known IOCs, and predefined rules. While effective for known threats, they struggle in modern environments.

Some Limitations:

- Detects only known threats
- Misses zero-day attacks
- Struggles with living-off-the-land techniques
- High false positives and alert fatigue

Dataset Overview

For this demonstration, we use a sample process execution dataset containing:

- Timestamp
- Host
- User
- Process Name
- Process Count/Frequency

The dataset simulates real enterprise telemetry where most processes are common, and only a few are truly suspicious outliers.

Why We Encode Process Names?

Machine learning and statistical algorithms work best with numerical data.

Challenges with raw process names:

- Text values cannot be directly used in many algorithms
- Difficult to compute statistical distance
- Hard to standardize behavior mathematically

Solution: Convert process names into numeric representations while preserving analytical value.

Numeric Transformation Approach

In this demo, process names are transformed into numeric values to enable statistical analysis.

Conceptual workflow:

1. Ingest process data into Splunk
2. Transform process names → numeric values
3. Normalize if required
4. Apply outlier detection logic
5. Identify rare processes

This enables Splunk to mathematically evaluate process behavior patterns.

Smart Outlier Detection in Splunk

- What looks different from normal?
- Which processes occur very rarely?
- Where should analysts investigate first?

In Splunk, this can be achieved using:

- Statistical commands
- Z-score–based detection
- Density-based approaches
- Frequency analysis

The focus of this demo is **rarity-based anomaly detection**.

Splunk MLTK Add-on Overview

The Machine Learning Toolkit (MLTK) is a powerful Splunk add-on that enables advanced statistical and machine learning workflows directly within SPL.

Key capabilities:

- Built-in ML algorithms
- Guided modeling interface
- Outlier detection techniques
- Forecasting and clustering
- Custom model creation

Let’s See the Demo

Any Questions?

Thank You

For Professional Red Team / Blue Team / Purple Team / Cloud Cyber Range labs / Trainings

Please Contact :

support@cyberwarfare.live

To know more about our offerings, please visit: **cyberwarfare.live**