Sentinel: Intelligent Cyber Threat Defender

CONTENTS

CONTENTS

01

Opening

AI-Powered Cybersecurity & Threat Intelligence System

Real-Time Risk Scoring, Threat Detection & Automated Insights

This presentation introduces an AI-powered cybersecurity system that offers real-time risk scoring, threat detection, and automated insights. It leverages advanced technology to provide comprehensive protection against cyber threats.

02

Project Overview

End-to-End AI Pipeline Overview

End-to-End AI Pipeline

The project features an end-to-end AI pipeline designed to detect cyber threats in real-time. It integrates multiple components for seamless threat detection and analysis.

Real-Time Ingestion and Analysis

Real-time data ingestion, machine learning anomaly detection, and large language model analysis are combined with interactive dashboards to provide actionable insights.

Phishing Detection Module

A specialized module for email, URL, and SMS phishing detection is included, enhancing the system's ability to identify and mitigate phishing threats.

03

System Architecture

High-Level System Architecture

Data Ingestion Layer

The data ingestion layer uses Kafka-style simulation for streaming data, ensuring continuous and real-time data flow into the system.

Spark Processing Engine

Spark processes the incoming data, cleaning,normalizing, and transforming logs to prepare them for analysis.

ML Anomaly Detection

Machine learning models detect anomalies and generate risk scores, providing a quantitative measure of potential threats.

Threat Classifier

A threat classifier using TF-IDF and Logistic Regression analyzes Gmail, URL, and SMS data to identify and score potential threats.

04

Data Flow

Data Ingestion & Processing Pipeline

Simulated Streaming

Simulated streaming using Python and Kafka-style events ensures a steady flow of data into the system for continuous monitoring.

Spark Data Processing

Spark cleans, normalizes, and transforms the data, maintaining module independence for efficient and scalable processing.

Intelligence Layer: ML, RL & LLM Engine

- ✓ The system integrates a multi-layer intelligence engine combining machine learning, reinforcement learning and large language models. Unsupervised anomaly detection is performed using Isolation Forest, while Random Forest Classifier identifies malicious vs. normal events based on aggregated Spark features from network logs, login activity and user behavior.

- ✓ A custom Gym cyber-environment then uses a PPO-based RL agent (with Q-learning fallback) to recommend mitigation actions such as blocking IPs, quarantining devices, throttling traffic or alerting administrators. This combined approach produces risk scores, policy-guided actions and incident classifications used across the SOC workflow.

05

ML Detection

ML Anomaly Detection Engine

01

Isolation Forest for Anomaly Detection

The Isolation Forest algorithm is used for anomaly detection, identifying unusual patterns in the data.

Risk Score Generation

Risk scores are generated by combining anomaly severity with classifier predictions, scaled to a 0–100 range for easy interpretation.

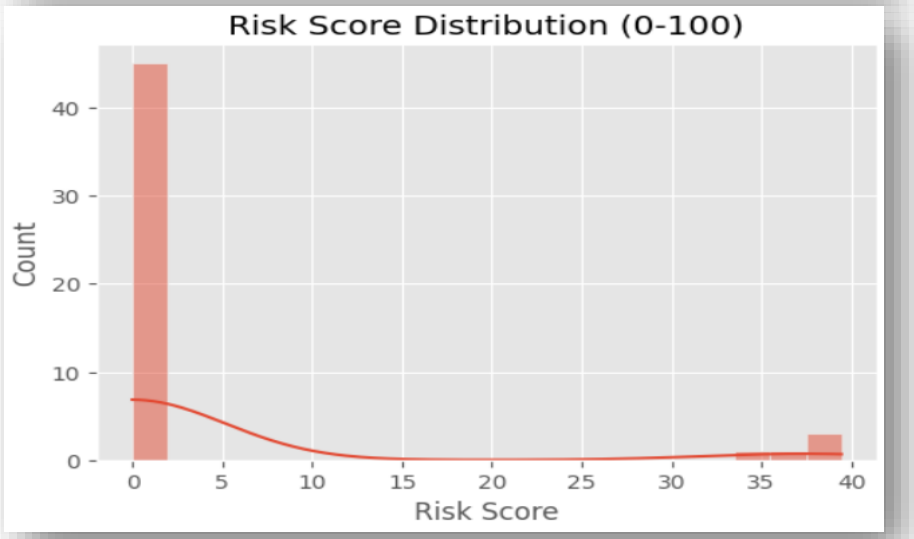
02

ML Anomaly Detection Engine

```
===== Part 4: ML Anomaly Detection Demo =====
[INFO] Anomaly detector trained and saved.
[INFO] Incident classifier trained and saved.

[Confusion Matrix]
[[45  0]
 [ 0  5]]

[Classification Report]
              precision    recall  f1-score   support

         0.0       1.00      1.00      1.00        45
         1.0       1.00      1.00      1.00         5

    accuracy                           1.00        50
   macro avg       1.00      1.00      1.00        50
weighted avg       1.00      1.00      1.00        50


[Sample Anomalies / Risk Scores]
Index: 1, Risk Score: 33.93, Label: 1.0, Anomaly Score: 0.06
Index: 2, Risk Score: 39.02, Label: 1.0, Anomaly Score: 0.13
Index: 15, Risk Score: 36.02, Label: 1.0, Anomaly Score: 0.09
Index: 17, Risk Score: 37.82, Label: 1.0, Anomaly Score: 0.11
Index: 27, Risk Score: 39.44, Label: 1.0, Anomaly Score: 0.13
```

✅ Model trained — phishing detection basic demo.

	precision	recall	f1-score	support
0	0.50	0.71	0.59	7
1	0.60	0.38	0.46	8
accuracy			0.53	15
macro avg	0.55	0.54	0.52	15
weighted avg	0.55	0.53	0.52	15



06

Dashboards

Risk Score & Mitigation Dashboard

Risk Score Trend

The Risk Score Trend chart helps analysts monitor the severity of incidents over time, providing a visual representation of threat levels.

Mitigation Cost Trend

The Mitigation Cost Trend chart tracks the financial impact of incidents, aiding in cost-effective threat management.

Dashboard Insights

These trends provide actionable insights for analysts to prioritize and address threats based on severity and financial impact.

Mitigation cost & actions

Mitigation Cost Trend

Final Results (Sample):

	timestamp	risk_score	mitigation_cost
0	2025-11-07 10:00:00	8	4871.93
1	2025-11-07 10:10:00	5	3697.00
2	2025-11-07 10:20:00	6	13394.11
3	2025-11-07 10:30:00	8	10213.38
4	2025-11-07 10:40:00	4	11496.87

Executive Summary

Executive Summary

Incident Type: Unauthorized Access Attempt Attempt . Risk Score: 8.5/10

Actionable Recommendations

1. Block suspicious IP
2. Reset credentials
3. Quarantine affected servers
4. Audit logs
5. Notify admin team

Dashboard Insights

AI Cybersecurity Dashboard

Type: login_fail **Risk:** 8/10 **Status:** Actioned **Action:** Block IP
Source IP: 192.168.0.4 **Asset:** Server-4

Events Table

	event_id	timestamp	event_type	source_ip	risk_score	status	applied_action
0	1	2025-11-07 10:00:00	cctv_motion	192.168.0.0	8	active	
1	2	2025-11-07 10:05:00	login_fail	192.168.0.1	5	active	
2	3	2025-11-07 10:10:00	cctv_motion	192.168.0.2	4	active	
3	4	2025-11-07 10:15:00	cctv_motion	192.168.0.3	8	active	
4	5	2025-11-07 10:20:00	login_fail	192.168.0.4	8	actioned	Block IP
5	6	2025-11-07 10:25:00	login_fail	192.168.0.5	3	active	
6	7	2025-11-07 10:30:00	cctv_motion	192.168.0.6	6	active	
7	8	2025-11-07 10:35:00	malware_alert	192.168.0.7	5	active	
8	9	2025-11-07 10:40:00	cctv_motion	192.168.0.8	2	active	
9	10	2025-11-07 10:45:00	cctv_motion	192.168.0.9	8	active	

✅ Action 'Block IP' applied to event 5.

07

Phishing Module

Gmail URL SMS Threat Detection

Simulated Email Inbox

A simulated inbox of 50 emails is processed using TF-IDF vectorizer and Logistic Regression to detect phishing threats.

Action Determination

The combined score determines the action: Quarantine, Review, or Safe, ensuring appropriate handling of potential threats.

URL and Attachment Analysis

URLs are extracted, and attachments are scored for risk, combining metrics to determine the overall threat level.

Risk Distribution

The module generates risk distribution and sender vs risk plots, providing detailed insights into threat sources.

08

Insights

Key Security Insights Uncovered

- The system provides key insights, including average risk levels, mitigation costs, phishing email percentages, and identification of risky senders and subjects.
- The system exports final results as structured CSV files, risk and cost trend plots, and an automatically generated PDF summary stored. These outputs compile end-to-end analytics from ingestion, detection, RL actions and business impact scoring.
- The module also includes deployment mapping for GCP, covering Pub/Sub, Big Query, Vertex AI and Cloud Run, enabling seamless migration to a production environment.

09

Tech Stack

Technology Stack Overview

Core Technologies

The technology stack includes Python, Spark, Kafka-style simulation, Scikit-learn models, TF-IDF, Pandas, NumPy, Matplotlib, and Plotly.

Optional Technologies

Optional technologies like YOLO and Stable-Baselines3 are considered for future enhancements, expanding the system's capabilities.

Dashboard Tools

Streamlit and Dash are used for creating interactive dashboards, providing real-time visualizations and insights.

10

Challenges

Deployment Challenges & Fixes

Ubuntu Repo Errors

Ubuntu repo 404 errors were resolved by rewriting the sources.list file, ensuring smooth package installation.

Java Gateway Stability

Java 11 was manually installed to stabilize the Spark gateway, ensuring reliable data processing and analysis.

11

Closing

Conclusion & Next Steps

Complete Cybersecurity Framework

The system delivers a complete real-time AI cybersecurity framework, integrating ML, NLP, analytics, and automated threat scoring.

Modular and Scalable Design

The modular and scalable design ensures the system can be easily extended and adapted for enterprise use.

Future Extensibility

With its robust foundation, the system is poised for future enhancements, ensuring continuous protection against evolving cyber threats.

YOUR LOGO

THANK YOU