Threat Intelligence Aggregator (Non-AI)
(Project Documentation)

1. Project Overview / Description

This project focuses on developing a practical Threat Intelligence (TI) Aggregator designed to collect, parse, normalize, and correlate threat intelligence indicators from multiple external and internal feeds — without using AI or machine learning.

Modern cybersecurity operations rely heavily on accurate and real-time threat intelligence. Indicators of Compromise (IOCs) such as malicious IPs, URLs, domains, and file hashes are distributed across multiple sources, often with different formats and inconsistent structures.

This project helps build a unified system that standardizes these feeds, correlates repeated indicators, and produces blocklists and actionable intelligence reports.

2. Practical Motivation

Organizations receive threat feeds from:

- Open-source intelligence platforms
- Commercial TI providers
- Security tools (SIEM, firewall, IDS logs)
- Government CERT notifications

Since formats vary (CSV, JSON, STIX, TXT, RSS), a normalized aggregator makes threat analysis faster and more reliable.

This project provides practical exposure to:

- How IOCs are structured
- How threat feeds are consumed
- How defenders correlate repeated or high-severity IOCs

- How automated blocklists are generated for security enforcement

3. Project Objectives

1. Collect threat intelligence from multiple TI feeds (files or URLs).

2. Normalize indicators into a unified format.

3. Parse IOC types: IPs, domains, URLs, hashes, emails.

4. Build a correlation engine to identify repeated or high-frequency indicators.

5. Generate blocklists for firewalls, IDS/IPS, and endpoint tools.

6. Export final reports and IOC datasets.

4. Practical Scope of the Project

A. IOC Feed Parser:

- Accepts multiple formats (CSV, TXT, JSON, STIX).

- Extracts IPs, URLs, domains, hashes, and email indicators.

- Removes duplicates and invalid entries.

B. Normalization Engine:

- Convert all indicators to a unified structure.

- Add metadata (source, timestamp, category).

C. IOC Correlation Engine:

- Identify indicators appearing across multiple feeds.

- Prioritize repeated indicators as high-risk.

- Generate severity ratings (Low/Medium/High).

D. Blocklist Generator:

- Create blocklists for:

  * Firewalls (IP set)

  * Web filters (malicious URLs & domains)

* EDR/AV (hash-based blocking)

- Support export formats (TXT, CSV, JSON).

E. Reporting Module:

- Summary of feeds processed

- Total unique indicators

- High-priority repeated indicators

- Exportable threat report

5. Tools & Technologies Used

Programming Languages:

- Python (recommended)

Libraries Used:

- re (regular expressions)

- json / csv (data parsing)

- requests (fetching feed URLs)

- ipaddress (IP validation)

- hashlib (for verifying hash formats)

Optional TI Formats:

- STIX/TAXII feeds

- Local IOC files

- Public OSINT sources

Documentation Tools:

- Word / Google Docs

- Draw.io for flowcharts and diagrams

6. Practical Techniques Implemented

Threat Intelligence Techniques:

- IOC parsing and validation

- Normalization of heterogeneous data

- Cross-source correlation

- Prioritization and enrichment

- Blocklist generation and export

Blue Team Defensive Techniques:

- Using blocklists for proactive defense

- Detecting repeated malicious infrastructure

- Enhancing SOC workflows with TI automation

7. Workflow / Architecture (Practical Explanation)

STEP 1: Load Feeds

- Import IOC feeds from URLs or files.

STEP 2: Parse Indicators

- Extract IPs, URLs, domains, hashes, emails.

STEP 3: Normalize Data

- Clean and standardize indicators.

STEP 4: Correlation Engine

- Find matches across multiple feeds.

- Identify high-priority repeated IOCs.

STEP 5: Blocklist Generation

- Create category-based blocklists.

STEP 6: Final Report

- Summaries, counts, categorized IOCs, risk scores.

8. Flowchart (Text Version)

START

↓

Load IOC Feeds

↓

Parse Indicators

↓

Normalize & Validate Data

↓

Correlation Engine (Cross-Feed Matching)

↓

Generate Blocklists

↓

Export Final TI Report

↓

END

9. Expected Practical Output

The final toolkit should output:

- Normalized IOC database

- Parsed list of IPs, domains, URLs, and hashes

- Correlation results (repeated indicators)

- Blocklists ready for deployment

- Final TI report summarizing all intelligence gathered

Expected Output Examples:

- High-risk IPs appearing in 5+ feeds

- Consolidated malicious domain list

- Hashes flagged across multiple TI sources

10. Learning Outcomes

This project enables understanding of:

- How threat intelligence feeds work

- IOC structures and validation

- Data normalization and parsing techniques

- Practical SOC and blue-team TI workflows

- How blocklists improve defensive posture

11. Project Deliverables

1. Project documentation (Word/PDF)

2. Working TI Aggregator toolkit

3. IOC correlation output

4. Blocklist files (IP/URL/hash)

5. Screenshots of modules running

6. Final presentation (PPT)