Threat Intelligence Feed Aggregator Documentation

July 13, 2025

Contents

1 Solution Overview

This document outlines the **AI-powered Threat Intelligence Feed Aggregator**, a platform designed for security teams, detection engineers, and researchers to efficiently monitor and respond to emerging cyber threats. The system aggregates, analyzes, and summarizes threat data from multiple online sources, providing actionable insights through an interactive interface.

2 Architecture & Workflow

2.1 Core Modules

The system comprises the following core modules:

Table 1: Core Modules of the Threat Intelligence Feed Aggregator

Module	Purpose
Feed Collector	Fetches and parses RSS/Atom feeds from curated threat intelligence sources.
IOC Extractor	Uses regex and web scraping to extract Indicators of Compromise (IOCs) such as IPs, URLs, and hashes from feed content.
LLM Summarizer	Integrates with local Large Language Models (via Ollama) to produce concise, actionable summaries.
Dashboard (UI)	Provides an interactive Gradio-based web interface for browsing, searching, and refreshing feeds.
Pipeline Orchestrator	Automates feed ingestion, IOC parsing, and summarization in a modular workflow.

2.2 Data Flow

The workflow follows these steps:

1. **Feed Ingestion:** Fetches posts from RSS/Atom feeds (e.g., security blogs, GitHub repositories) using feedparser.

2. **IOC Extraction:** Applies regex patterns to extract IOCs (e.g., IP addresses, URLs, file hashes) from raw content.

3. **AI Summarization:** Sends parsed articles to a local LLM (e.g., LLaMA 2 via Ollama) to generate concise summaries highlighting threat type, affected entities, and recommended actions.

4. **Dashboard Presentation:** Displays feeds, extracted IOCs, and AI summaries in an interactive Gradio dashboard.

5. **User Interaction:** Enables searching, filtering, and manual refresh of feeds and summaries.

3 Technology Stack

The system leverages the following technologies:

Table 2: Technology Stack

Component	Technology/Library
Backend	Python
Feed Parsing	feedparser
IOC Extraction	re (regex), BeautifulSoup (optional for HTML parsing)
AI Summarization	Ollama (local LLMs: LLaMA 2, GPT-J, etc.)
Web Interface	Gradio
Pipeline	Custom Python scripts or pipeline libraries
Storage	CSV/JSON files or lightweight database (optional)

4 Key Features

- Aggregates threat intelligence from multiple curated feeds.
- Automatically extracts IOCs (IPs, URLs, hashes) from content.
- Generates AI-powered summaries for rapid understanding.
- Provides an interactive dashboard with search, browse, and refresh capabilities.
- Features a simple, modular, and well-documented codebase.

5 Deployment & Usage

To deploy and use the Threat Intelligence Feed Aggregator:

1. **Install Dependencies:** Run `pip install feedparser gradio requests` to install required Python libraries.

2. **Install Ollama:** Set up Ollama and run a local LLM model (e.g., LLaMA 2).

3. **Configure Feed URLs:** Specify the RSS/Atom feed URLs in the configuration file.
4. **Run the Pipeline:** Execute the main Python script to start the data processing pipeline.
5. **Launch Dashboard:** Start the Gradio dashboard for interactive access.

6 Example Use Cases

- **SOC Dashboard:** Offers a centralized view of the latest threat reports and IOCs for incident response.
- **Research Platform:** Provides aggregated, searchable threat data for security analysts.
- **Rapid Triage:** Accelerates investigation and response with AI summaries and IOC extraction.

7 Contact Information

For further details or support, please refer to the project repository or contact the development team through the official project channels.