ADAPTIVE THREAT INTELLIGENCE: AN AI-INTEGRATED HOLISTIC FRAMEWORK FOR MODERN CYBERSECURITY CHALLENGES

ABSTRACT

Modern cybersecurity landscapes are characterized by an ever-increasing volume and sophistication of cyber-attacks, necessitating proactive and intelligent defense mechanisms. Cyber Threat Intelligence (CTI) has emerged as a critical asset for organizations to understand and mitigate these threats. However, many existing approaches to CTI management are fragmented, focusing on isolated aspects like gathering or storage without comprehensive, AI-driven enrichment and correlation. This paper introduces ThreatWise AI, a novel, holistic framework designed to automate and enhance the entire CTI lifecycle. The proposed framework integrates advanced tools for gathering data from diverse internal (e.g., honeypots, system logs) and external sources (e.g., surface, deep, and dark web, social media). It leverages Artificial Intelligence (AI) and Machine Learning (ML) for critical tasks such as data classification, Named Entity Recognition (NER), outlier detection, and advanced correlation of threats. By utilizing a customized MISP platform for storage, enrichment, and sharing, ThreatWise AI provides a streamlined, efficient, and secure process for generating actionable intelligence. Experimental results on real-world data demonstrate the framework's effectiveness in classifying cybersecurity content, identifying novel attack patterns, and correlating disparate threat information, thereby offering a significant advancement over conventional CTI management systems.

EXISTING SYSTEM

The prevailing paradigm in CTI management often relies on a collection of disparate tools and platforms that address individual stages of the intelligence lifecycle in isolation.

Main Disadvantages of the Existing System:

1. Fragmented and Non-Integrated Architecture: Many systems specialize in either internal monitoring (e.g., using honeypots with ELK stack) or external data collection (e.g., using

standalone crawlers), but lack a unified architecture that seamlessly correlates intelligence from both streams. This leads to a siloed view of the threat landscape.

2. Limited Intelligent Enrichment: A significant shortcoming is the reliance on basic, rule-based parsing and correlation. These systems often lack advanced AI/ML modules for tasks like contextual classification, NER for extracting cybersecurity-specific entities, and sophisticated outlier detection to distinguish novel attacks from common noise.
3. Inflexible and Superficial Data Gathering: Existing crawlers often fail to adapt to modern web defenses. They cannot handle dynamically loaded content or evade anti-bot measures, resulting in incomplete data collection from critical sources like social media or dark web forums. Furthermore, classification is often binary (relevant/irrelevant) and does not extend to domain-specific categorization (e.g., aviation, naval), limiting the contextual relevance of the intelligence.

PROPOSED SYSTEM

The proposed system, ThreatWise AI, is a holistic framework that integrates all aspects of CTI management into a cohesive, AI-driven pipeline.

Main Advantages of the Proposed System:

1. Unified Holistic Architecture for End-to-End Management: ThreatWise AI provides a single, integrated framework that combines data gathering from diverse internal sources (honeypots, Wazuh agents) and external sources (web, social media) with advanced AI-powered analysis and the MISP platform for storage and sharing. This ensures a continuous and automated CTI lifecycle.
2. Advanced AI/ML for Deep Enrichment and Actionable Insights: The framework incorporates multiple sophisticated AI modules. This includes a Focused Crawler with a cybersecurity classifier, a domain classification model combining text and stylometric features, transformer-based NER (BERT, XLNet), an outlier detection module using Isolation Forest to identify novel attacks, and advanced correlation techniques using ARL and text similarity. This multi-layered enrichment significantly increases the quality and actionability of the CTI.

3. Adaptive and Evasive Data Collection Capabilities: The framework features an Evasive Crawler that mimics human behavior using tools like Selenium, allowing it to bypass anti-bot measures and access dynamically loaded content from surface and dark web sources. This, combined with its ability to perform domain-specific classification, ensures the collection of more complete and contextually relevant threat data that is missed by conventional systems.

SYSTEM REQUIREMENTS

- **H/W System Configuration:-**
- Processor - Pentium –IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB
- Key Board - Standard Windows Keyboard
- Mouse - Two or Three Button Mouse
- Monitor - SVGA

SOFTWARE REQUIREMENTS:

- **Operating system** : Windows 7 Ultimate.
- **Coding Language** : Python.
- **Front-End** : Python.
- **Back-End** : Django-ORM
- **Designing** : Html, css, javascript.
- **Data Base** : MySQL (WAMP Server).