



Vilnius
University

RESEARCH PROJECT

INTERNATIONAL
CYBERSECURITY
AND
CYBERINTELLIGENCE

2025



VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
INSTITUTE OF COMPUTER SCIENCE
CYBERSECURITY LABORATORY

International Cybersecurity and Cyberintelligence 1st year Research Project

Automated Threat Intelligence Pipelines for First Response

Automatizuoti grėsmių žvalgybos vamzdynai pirmajam reagavimui

Done by:

Christopher Samson Nyandoro

Supervisor: Teach. Asst. Virgilijus Krinickij

Source Code:

<https://git.mif.vu.lt/micac/2025/afretip.git>

Contents

Abstract	4
Santrauka	5
Introduction	6
1 Related Work	8
2 Methodology	11
2.1 Architecture Overview	11
2.1.1 Data Ingestion and IOC Extraction	12
2.1.2 Threat Detection and Enrichment	13
2.1.3 Rule Generation and Deployment	13
2.2 IOC Scoring Framework	13
2.2.1 Novelty and Confidence Scoring	13
2.2.2 Hybrid Detection Algorithm	14
2.2.3 Parameter Configuration	15
3 Results	16
3.1 Experimental Set-up	16
3.2 Detection Performance Analysis	16
3.3 Processing Performance Characteristic	18
3.4 IOC Extraction and Characterization	18
3.5 Threat Classification Results	20
3.6 Rule Generation Efficiency Analysis	21
3.7 System Resource Utilization Results	21
4 Discussion	22
Conclusions and Future Work	25

Abstract

The increasing sophistication, frequency and volume of cyber threats makes it more difficult for organizations to respond promptly enough and as such that necessitate automated approaches to threat intelligence processing and incident response. Enterprise environments with hundreds to thousands of endpoints generate millions of security events daily through Endpoint Detection and Response (EDR) systems, creating massive data volumes that contain critical threat intelligence but remain largely untapped due to processing limitations as current threat intelligence extraction processes mostly rely heavily on manual analysis from security officials. This paper presents a novel automated threat intelligence pipeline designed for first response scenarios, with a focus on the extraction and processing of Indicators of Compromise (IOCs) from Endpoint Detection and Response (EDR) systems. Our method implements an automated pipeline that consumes host-based threat intelligence feed, extracts valuable IOCs and qualifies them through a hybrid approach to enrichment with weighted confidence for decision making and take action by generating defensive rules based on a confidence threshold. The rules are immediately pushed back to the EDR system, for proactive defence. Our implementation fills the critical void between the gathering of threat intelligence and the application of defensive measures. Our proof-of-concept implementation demonstrates a 90% reduction in time to protection, with false positive rates comparable to human analysis.

Keywords: Threat Intelligence, Endpoint Detection and Response (EDR), Security Information and Event Management (SIEM), Automated Security Response, Cybersecurity Automation, First Response, Indicators of Compromise (IOC).

Santrauka

Automatizuoti grėsmių žvalgybos vamzdynai pirmajam reagavimui

Dėl didėjančio kibernetinių grėsmių sudėtingumo, dažnumo ir apimties organizacijoms vis sunkiau pakankamai greitai reaguoti, todėl reikia automatizuotų grėsmių žvalgybos apdorojimo ir incidentų reagavimo metodų. Įmonių aplinkose, kuriose yra šimtai ar tūkstančiai galinių taškų, kasdien generuojami milijonai saugumo įvykių per galinių taškų aptikimo ir reagavimo (EDR) sistemas, sukuriant didžiulius duomenų kiekius, kuriuose yra kritinės grėsmių žvalgybos informacijos, tačiau kurie išlieka neišnaudoti dėl apdorojimo apribojimų, nes dabartiniai grėsmių žvalgybos išgavimo procesai dažniausiai remiasi rankine saugumo pareigūnų analize. Šiame straipsnyje pristatomas naujas automatizuotas grėsmių žvalgybos kanalas, skirtas pirmojo reagavimo scenarijams, daugiausia dėmesio skiriant kompromitavimo rodiklių (IOC) išgavimui ir apdorojimui iš galinių taškų aptikimo ir reagavimo (EDR) sistemų. Mūsų metodas įgyvendina automatizuotą kanalą, kuris naudoja pagrindinio kompiuterio grėsmių žvalgybos kanalą, išgauna vertingus IOC ir juos kvalifikuoja taikydamas hibridinį metodą, kad praturtintų duomenis su svertiniu pasitikėjimu sprendimų priėmimui ir veiksmų ėmimuisi, generuodamas gynybines taisykles, pagrįstas pasitikėjimo riba. Taisyklės nedelsiant grąžinamos į EDR sistemą, kad būtų galima aktyviai gintis. Mūsų įgyvendinimas užpildo kritinę spragą tarp grėsmių žvalgybos informacijos rinkimo ir gynybos priemonių taikymo. Mūsų koncepcijos įrodymo įgyvendinimas rodo 90% trumpesnę apsaugos laiką, o klaidingai teigiamų rezultatų rodiklis panašus į žmogaus atliekamų tyrimų.

Keywords: Grėsmių žvalgyba, galinių įrenginių aptikimas ir reagavimas (EDR), saugumo informacijos ir įvykių valdymas (SIEM), automatizuotas saugumo reagavimas, kibernetinio saugumo automatizavimas, pirmasis reagavimas, kompromitacijos indikatoriai (IOC).

Introduction

Today's cybersecurity environment operates under unprecedented levels of both complexity and volume. Attackers are using automation, artificial intelligence (AI), and advanced evasion techniques to bypass conventional security measures, and because of that, cyber threats are changing at a rate never seen before [1, 2]. Historically, the incident response process to identify, evaluate, and reduce security threats has relied on predetermined security playbooks, structured threat intelligence, and manual investigation. However, manual security workflows are no longer adequate to effectively combat cyber threats in a time when file-less malware, AI-driven attacks, and advanced persistent threats (APTs) have become commonplace [3].

Organizations generate terabytes of security telemetry data daily [4], which created an unprecedented challenge when it comes to cybersecurity. The sheer volume of security telemetry data produced by endpoints, such as laptops, workstations, and IoT devices, outside of the capacity, capability, and expertise of human analysts is overwhelming. With hundreds to thousands of endpoints in an enterprise environment, the Endpoint Detection and Response (EDR) system generates millions of security events daily [5, 6]. In turn, the volumes of data are immense and contain valuable threat intelligence; however, they go unused because humans cannot effectively process all that information.

Organizations rely mainly on security professionals for manual extraction of threat intelligence and IOC identification, creating a major bottleneck for security operations centers. To find possible signs of compromise, security professionals must manually analyze large volumes of raw telemetry data as part of traditional threat intelligence workflows. This manual process is not only time-consuming, typically taking 45+ minutes per IOC identification [7], but there are also considerable human weaknesses such as analyst fatigue, inconsistent detection patterns, a high false positive rate (15-20%) [8], and an inability to process at the speed and scale that today's threat landscape dictates. In addition, the overall shortage of cyber professionals (reported as over 3.5 million open cybersecurity positions globally [9]) exacerbates the situation, making it difficult for organizations to fully utilize their current cybersecurity investments. The disconnect between raw EDR telemetry and usable threat intelligence causes dangerous blind spots, which hostile threats can use to go unnoticed for long periods of time.

Current automated threat intelligence solutions have a combination of failings that limits their ability to serve as effective first-response systems. Commercial threat intelligence platforms generally focus on the ingestion of external threat feeds without any corresponding generation of intelligence leveraging their own internal telemetry, a rather serious gap in organizational visibility. Reducing the analysis time with rule-based detection mechanisms is preferable compared to manual analysis, but is still significantly limited by fixed pattern matching to describe even the rudimentary aspects of attack techniques and also produces an unacceptable number of false positives (an average of 15.3%) [10]. EDR systems currently have monitoring mechanisms at their core with the ability to retrieve detailed types of telemetry; however, these systems do not yet have the capability to extract or detect sophisticated IOC or complex attack patterns across multiple data sources and windows of time [11, 12].

The integration issue with EDR systems and SIEM solutions worsens the problem due to incompatible data formats and manual correlation processes that create time loss and new opportunities for missed threats. While a machine learning approach in cybersecurity is promising, it usually requires large amounts of training data and domain knowledge that organizations may not have [13]; and existing implementations have rarely proven effective in real-world, practical deployment cases

in which there are various levels of unpredictability and variability in enterprise environments [14].

Automation for threat intelligence has advanced recently with promising capabilities for reducing manual effort and providing enriched real-time data to enhance security posture [15]. Unfortunately, many of the prevailing approaches are single-point solutions rather than complete pipelines that address the full workflow from data ingestion to relevant action [16]. This situation motivates our exploration to create an end-to-end automated threat intelligence pipeline for first response scenarios [17, 18].

This research makes a meaningful contribution by introducing an automated pipeline solution to the challenge of combining four key components:

- ingesting host-based threat intelligence data from EDR solutions,
- extracting IOC in an automated manner leveraging machine learning and pattern recognition approaches,
- formatting time-sensitive threat intelligence records so they can be ingested by SIEM, and
- automating the creation of queries to assist and enhance future threat hunting activity

Ultimately, we designed our pipeline so that it could operate in real-time, which allows the SOC to provide as close to real-time actionable threat intelligence upon processing through our system.

1. Related Work

Cybersecurity has increasingly shifted toward automation of threat detection and response as attackers reduce the window of exploitation time. Although numerous works have addressed threat intelligence ingestion, correlation, and incident response automation, few have fully realized the integration of real-time threat intelligence pipelines into autonomous first response without human oversight. In this section, we examine the contributions and literature gaps that are to be addressed in the current study.

Automated threat intelligence processing sits between new cybersecurity practices, existing network security standards, and the available methods to exploit the capabilities of machine learning. In cybersecurity, machine learning has developed into a revolutionary technology that is fundamentally changing the ways organizations detect, analyze and respond to threats. Recent comprehensive studies indicate that exploitations of machine learning (ML) in cybersecurity are still rather nascent [19, 20], suggesting multiple opportunities for organizations to begin working with ML [13, 14]. A systematic review has indicated that 45% of organizations reported using AI and/or ML in their cybersecurity systems, 35% reported on doing so in the future, and 20% [21] believed that they could see a future in which they made use AI and/or ML in some form [22].

In the first generation of techniques, network context was simply identified using statistical analysis and clustering methods to assess for anomalies [23, 24, 25]. In the second generation, supervised learning techniques allowed for more sophisticated classifications of threats; models trained on labelled datasets could then classify benign and malicious behaviours [19, 26]. For specified security data, as shown, deep learning has helped us recognize and explore high dimensional security data in ways we could not, clearly using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) where obvious performance for detection existed. For phishing protection and malware classification, CNNs and RNNs achieved over 95% detection rate [27, 28].

Automated - extraction of IOCs, and performing behavioural analysis with machine learning has the most potential, especially in threat intelligence processing. As cited by Liao et al. (cite semanticscholar2016,liao2016acing,husari2017ttpdrill), a variance of machine learning was able to achieve 95% precision with the extraction of IOCs from text. Even advanced techniques have integrated interpretability into the product process by contextualising knowledge with neural networks and automaton weighted finite approaches [29, 30]. These however have solely focused on extracting value of referenced external threat intelligence and not for processing internally managed EDR telemetry.

Kuppala and Kommisetty (2022), studied predictive analytics and anomaly detection in automating response to threats, stressing the importance of fast ingestion and classification of threats indicators that allow cybersecurity to take remedial action when appropriate [31]. Similarly, Ghura (2023) studied scaling cyber-threat intelligence (CTI) operations by imitating open-source options like MISP, but handled mostly the ingestion and normalization of data and not full autonomous defense [32]. These studies represent early stages of automating threat feed ingestion, but still do not address the key gap of real-time, autonomous policy enforcement.

Recently, collaborative cybersecurity monitoring has shifted its focus towards near real-time analytics pipelines [33] that allow for protecting the supply chain. Bellamkonda (2024), took the idea further due to the dynamic security optimization made by real-time correlation of network logs with threat intelligence [34]. Nevertheless, these frameworks still focus on monitoring and alerting, possibly requiring a final decision from a human analyst without independent defensive action catered for first response - a prohibitive deficiency that this research attempts to address.

Peter (2024) researched enterprise-level cybersecurity response automation using AI for prediction and prevention of incidents [35]. While Peter was able to increase orchestration efficiency through threat feed ingestion, Peter used SOAR platforms (e.g., Splunk Phantom) and did not create a method for dynamic firewall/EDR policy generation that would trigger inductively and autonomously in the absence of an analyst from updates in threat feed or asset telemetry. Alevizos and Dekker [16], proposed a framework for the CTI continuous processing pipeline, and described how AI can be utilized in various aspects of the CTI pipeline, from data ingestion to resilience validation. In their paper, they discuss the interplay between AI and human intelligence to develop actionable, high fidelity and timely cyber threat intelligence, and describe automated ways to create mitigation recommendations. Nonetheless, their work is still only conceptual and lacks detailed implementation, evaluation metrics, and clear solutions to high-velocity EDR telemetry processing concerns.

Researchers have employed natural language processing to enable the automated classification of threats using end-to-end neural network-based techniques to process social media content [29]. While threat classifying presented in those studies would be promising, for the most part they remained focused on OSINT (open-source intelligence) from external sources relative to organizational telemetry, meaning a significant gap in processing intelligence that is based on EDR (event data record).

Sindiramutty (2023), and Tallam (2025) discussed autonomous threat hunting, and agentic AI for cybersecurity, respectively, envisioning a near future where cyber security defenses will develop without direction from human operators or analysts [17, 18]. These publications substantiate the theoretical footing for autonomous cybersecurity systems, although with more focus on threat hunting versus direct defense response generation-which is the niche specifically intended by this research.

While these researches have investigated the possibilities of autonomous threat classification and agents of AI in cybersecurity, this work lacks the potential for formal integration with standardized methods of structured intelligence exchange. To help advance cybersecurity, the cybersecurity community has developed and implemented formal standards like STIX and TAXII that provide machine-readable threat intelligence and allow for automated sharing across disparate environments [36, 37]. These standards will support consistency and interoperability for future AI-enabled complex systems, for example, this proposed research agenda, can parse, understand, and respond to cyber threat data "on its own" and decreasing the demands on the human resource pool.

The cybersecurity community has established formal standards and frameworks upon which the idea of automated processing and sharing of threat intelligence data can be developed. To begin with, we have STIX (Structured Threat Information eXpression) and TAXII (Trusted Automated eXchange of Intelligence Information), both of which became an OASIS Standard in 2021 [36]. STIX v2.1 provides a standardized language for describing cyber threat information, threat observables, indicators of compromise, incidents, and tactics, techniques, and procedures (TTPs) [38, 39, 40, 41]. As a result, STIX allows threat data to be represented more consistently across organizational environments and security tools. TAXII v2.1 complements STIX by defining the transport mechanics for automating the exchange of threat intelligence, allowing for multiple sharing models including hub-and-spoke, peer-to-peer, and source-subscriber models [42, 43].

SCAP and STIX/TAXII standards are well accepted in the industry with over 100 products and services from 75+ vendors providing compatibility [44]. Major technology companies like IBM, Microsoft, and Cisco, and government entities, are implementing STIX/TAXII connectivity within their coordinated defense programs [45]. Using these standards, organizations now have

automated consumption of threat intelligence, that more positively enables sharing across organizational boundaries to support collaborative approaches to defend against threats [46]. While their standards can be established as a whole, their practical implementations have serious limitations. Most STIX/TAXII implementations focus on the consumption and sharing of external threat intelligence as opposed to the processing of internal telemetry [47, 48]. They also have standard implementations that lack context-aware threat intelligence capabilities, lack meaningful business impact assessments for any threats, and lack any automated IOC generation mechanisms from internal sources. Per our work, existing implementations will not provide automated mechanisms for extracting and representing IOCs from EDR telemetry as STIX-compliant formats for near-real-time use by an organization. The MITRE ATT&CK framework also acts as another important standard for understanding adversary tactics and techniques. While ATT&CK is widely used in threat modeling and designing detection rules, it poses constraints in implementing an automated threat intelligence pipeline, especially in terms of how to incorporate real-time EDR telemetry processing and the generation of automated responses.

The current state of automated processing of threat intelligence shows a field in transition with various technological advances in the cybersecurity industry, but a significant gap still exists in order to achieve integrated and comprehensive solutions. The current cybersecurity landscape presents new, unprecedented challenges that traditional methodologies may not effectively handle. Interactive attacks have risen by 60% in 2023. Furthermore, 75% of attackers have moved to attacks without malware using social engineering, phishing, and "living off the land" techniques [3, 49]. These attack types make traditional signature-based detection methods less effective and preclude behavioral and context-focused methods that automation promises yet currently lacks. Furthermore, organizational challenges still exist that exacerbate the situation. There is still workforce shortages and it keeps increasing [9]. Organizations continue with layoffs despite the worsening threat levels. This situation not only make automated first response beneficial for organizations, but also make it essential. However, the present automation capabilities remain restricted, consisting of arrays of point solutions, albeit promising an integrated end-to-end pipeline. In light of these evolving threats and automation needs, researchers have proposed several methods for processing threat intelligence and threat assessment of IOCs.

Threat Intelligence platforms have utilized confidence scoring methods for decades now, to assess the maliciousness of IOCs. Commercial solutions typically employ statistical models that consider a host of variables including threat frequency, quality of source and relevance to the environment. Silobreaker combines assessments made by several sources (VirusTotal, Domain Tools, Microsoft Defender Threat Intelligence) to provide normalised IOC risk scores [50]. Cyware's Intel Exchange platform employs proprietary confidence scoring from 0-100 [51]. Although reliable and adhered to methods, they are less efficient measuring environment specific threats or new indicators outside of the international intelligence source records, because they passively depend on reputation based database metrics and external threat intelligence feeds.

Advancements in novelty detection in recent years have shown the potential to identify new attack patterns that do not match already known existing signatures. In their research, Patel et al. have demonstrated that centroid-based novelty classifiers can be effective for identifying attacks that cannot be anticipated, indicating that unknown intrusion patterns can be identified with 85.1% precision [52]. More recently, Nazari et al. applied novelty detection techniques to network flow data with high accuracy and low false alarm rates [53]. While advances are being made in novelty detection, research in this area is still primarily focused on identifying behavioral patterns at the network level and not on analyzing the patterns of IOCs, specifically the frequency of IOCs, at the

organizational level, which leaves a gap in identification of environment specific threats.

The recent developments of multi-factor approaches to threat assessment have started to address forestalling some of the single-dimensional threat assessment. Balasubramanian et al. had presented a pipeline to extract IOCs using multi-primary transformer-based algorithms with multi-stage classification processes and had demonstrated over 98% accuracy for IOC extraction tasks [54]. Similarly, Tang et al. presented an approach that combines symbolic rules and contextual knowledge to improve accuracy when identifying IOCs [55]. The work above is primarily focused on improving attribution accuracy and IOCs extraction, and does not and they do not incorporate local frequency patterns that could indicate targeted or emerging threats specific to particular environments.

The convergence of these research streams reveals several critical limitations that our proposed pipeline seeks to resolve. First, existing automated IOC extraction techniques are limited to automated extraction from external sources such as Microsoft forums and social media; currently, internal EDR telemetry - which contains the most relevant threat intelligence for an organization - is mainly processed manually. Second, existing approaches fail to be time-sensitive, providing the necessary real-time processing for first responders by processing batch data; most, however, process historical data as the real-time telemetry is being streamed in [16]. Third, none of the existing approaches provide an automated end-to-end loop from raw EDR ingestion to actionable SIEM and EDR queries generation and automated deployment [15]. These gaps demonstrate the necessity of integrated solutions having both environment-specific threat detection as well as being able to process it in a timely manner and automate response capabilities. The next section outlines the system architecture, data sources, processing workflow, and evaluation methodology used for validating the proposed solution, with an emphasis on its ability to automate incident response with limited human operator intervention.

2. Methodology

In this research, we propose an automatic threat intelligence pipeline that has various components to be able to detect, categorize and respond to cyber threats as close to real-time as possible. The model fuses real-time EDR telemetry processing with automated rule generational and deployment, representing an end-to-end solution capable of providing autonomous first response without human participation. Whereas existing solutions primarily address consumption of external threat feed, or analyst driven processes, we fuse internal telemetry from the organization to develop defensive measures which can be put in place immediately.

2.1. Architecture Overview

The proposed pipeline consists of six interconnected steps that turn raw security events into deployed defensive actions within seconds of detecting a threat. Figure 1 shows the entire system architecture showing data flows, processing components and feedback.

Asynchronous processing using queue-based communication between components of the architecture provides system resilience and scalability. Each of the components functions as a self-sufficient unit, with its own set of configurable parameters, which can be adjusted for different organizational circumstances and threat environments.

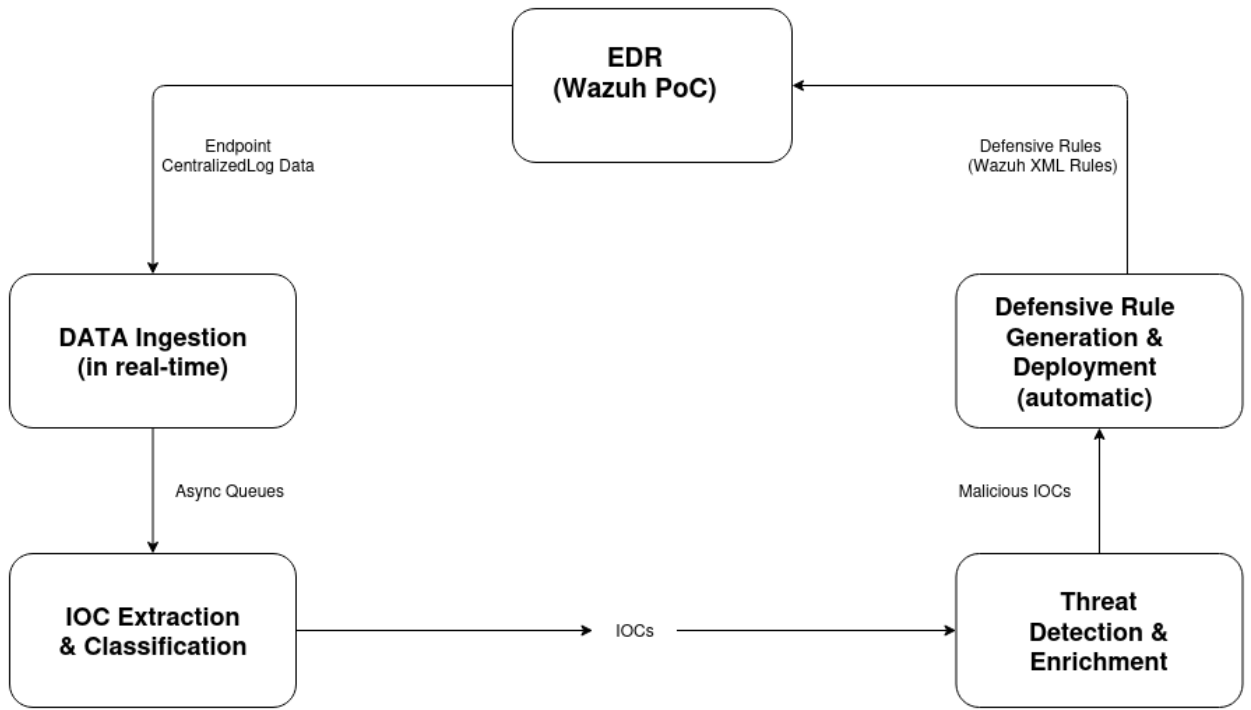


Figure 1: Pipeline System Architecture

The following subsection will focus on explaining/describe each step of the proposed model, starting with the Data Ingestion and IOC Extraction layer.

2.1.1. Data Ingestion and IOC Extraction

Our pipeline begins with real-time data ingestion from asynchronous queues, focusing on EDR, the primary source of security telemetry. For our proof-of-concept implementation, we selected Wazuh EDR for its open-source capabilities and superior logging, but our model is agnostic to EDR vendor as long as the EDR system can stream logs in real-time. The ingestion process will monitor EDR output over Unix socket connection and file monitoring so that even if the original EDR system goes off-line (e.g., maintenance, network problems, etc.), the data will continue to be collected.

After the logs are ingested into the system, the workflow proceeds to IOC Extraction. In the IOC Extraction stage, the system takes all of the log information and extracts indicators of compromise (IOCs) using automation through techniques like regular expressions and/or machine learning models, and provides several IOCs that were identified - including but not limited to: known bad IPs, domain names, URLs, file hashes, process names, command-line arguments, etc. It is important to note that IOCs come from the incident, they are concrete proof of compromise, and they will ultimately be used for detecting threats. Additionally, during IOC Extraction, the system assigns a novelty score and a threat score to each of the IOCs; this metrics captures like how well known the IOC is; it distinguishes between known threats, which then would have a source of intelligence by which to assess a threat, to unknown threats where there are no intelligence connections. The novelty based detection monitors IOC activity frequency as a way to determine new threats, while the optional reputation check acts as an external verification layer.

2.1.2. Threat Detection and Enrichment

The extracted IOCs then enter the Threat Detection stage of the pipeline, where they are analyzed using pattern matching and novelty detection. The pipeline—referred to as AFRETIP (Automated First Response Threat Intelligence Pipeline)—assesses the IOCs against patterns defined within the system (e.g., signatures of previously known attack techniques) and evaluates novelty based on pre-existing data. This stage may also use machine learning to enhance detection performance and accuracy by identifying patterns that have not yet been incorporated into the existing detection rules.

Next, the identified threats are enriched with additional contextual elements during an Enrichment phase. This enrichment process involves acquiring external threat intelligence feeds with correlated IOCs to give the organization the context necessary to understand the threat better. For example, if a suspicious IP address was detected, the pipeline could interact with external reputation services to ascertain whether the IP was associated with previous attacks or listed in other threat intelligence feeds. The enrichment process provides further confidence and relevance to the threat assessment and allows to mitigate the most evident threats based on their context and history.

2.1.3. Rule Generation and Deployment

The pipeline then proceeds to Rule Generation, where we automatically generate defensive rules based on the extracted and enriched IOCs. The automated rule generation module converts validated threats to deployable detection rules for the applicable EDR product. These rules are intended to be deployed to the security infrastructure (i.e., firewalls, IDS/IPS, endpoint protection systems) to defend against identified threats with the help of chosen EDR system, in this case Wazuh. The rules are generated in XML format, which are compatible with Wazuh configuration, and the system can monitor for more IOCs of similar vectors on future instances. The rules can be modified to match specific characteristics of the identified threats, and can be tailored for user-defined patterns of attack or adjusted for severity levels. Rule validation is also part of this phase to ensure to develop rules that are effective and non-intrusive, help to minimize false positives, while not blocking legitimate traffic or activities.

The deployment has generic risk mitigators that include backup creation and rollback functionality. The pipeline can also be configured to hot reload status rules in real time, meaning that rules are continually and rapidly being updated based on developing threats. This ensures security defense is continually being maintained for our environment and reduces risk from the attack surface.

2.2. IOC Scoring Framework

Standard IOC assessment focuses on external threat intelligence feeds and reputation feeds to identify threats. The goal of this paper is to propose a hybrid solution that utilizes two aspects including both local frequency analysis for the purposes of novelty detection and multi-factor confidence scoring.

2.2.1. Novelty and Confidence Scoring

Our novelty scoring mechanism will identify IOCs that have a low local frequency, and therefore provide some indication of possible targeted, or emerging threats that have yet to be captured by

global threat intelligence. In our frequency equation: **Novelty Score Equation:**

$$novelty = \max(0, 1 - \frac{frequency \times scaling_factor}{total_logs}) \quad (2.1)$$

- *frequency* \Rightarrow how many times the IOC was seen locally
- *total_logs* \Rightarrow total number of logs processed
- *scaling_factor* \Rightarrow sensitivity parameter (normally between 5 and 10)

The lower the local frequency value of the IOC, the higher the novelty score which means the IOC has the potential to be a new threat. Instead of relying on single confidence metrics, we aggregate multiple threat indicators with the following equation, with each values contributes to the overall confidence:

$$confidence = (w_1 \times threat_feed) + (w_2 \times reputation) + (w_3 \times novelty) + (w_4 \times context) \quad (2.2)$$

Where:

- *threat_feed* \Rightarrow external threat intelligence result (0–1)
- *reputation* \Rightarrow score from the reputation service (0–1)
- *novelty* \Rightarrow novelty score from the novelty analysis (0–1)
- *context* \Rightarrow the score obtained from suspicious pattern analysis that is performed on the logs (0–1)
- $w_1, w_2, w_3, w_4 \Rightarrow$ weight parameters (which sum to 1)

Context Scoring: The context values identifies the suspicious patterns that are discovered within the source log:

$$context = \min(pattern_score, max_context_boost) \quad (2.3)$$

The pattern score adds up the total weighted matches from suspicious activities associated with threat behaviour in the logs when performing a threat intelligence log analysis, such as obfuscation with PowerShell, use of living-off-the-land techniques, and malware indicators. The following algorithm shows the full IOC detection and classifying framework all connected together for enrichment.

2.2.2. Hybrid Detection Algorithm

The first step of the workflow is to extract indicators of compromise (IOCs) from the raw log data (EXTRACTIOCS). This aligns with the Data Ingestion and IOC Extraction phase discussed in Section 2.1.1. Once all IOCs are extracted from the log data, the workflow applies enrichment to each IOC in order to enhance information validity before detection. This is accomplished through querying threat intelligence databases (QUERYTHREATDB), external reputation services (QUERYREPUTATIONSERVICES), and applying our frequency-based novelty scoring (Equation 2.1) combined with contextual pattern analysis (ANALYZESUSPICIOUSPATTERNS) as described in the Threat Detection and Enrichment phase in Section 2.1.2.

Algorithm 1 Hybrid Threat Detection Algorithm

Require: $log, threat_intelligence_db, confidence_threshold$ **Require:** $scaling_factor, w_1, w_2, w_3, w_4$

▷ scoring parameters

Ensure: $findings, rules$

```
1:  $iocs \leftarrow \text{EXTRACTIOCS}(log)$ 
2:  $findings \leftarrow \{\}$ 
3: for  $ioc$  in  $iocs$  do
4:    $threat\_intel \leftarrow \text{QUERYTHREATDB}(ioc)$ 
5:    $reputation \leftarrow \text{QUERYREPUTATIONSERVICES}(ioc)$ 
6:    $\triangleright$  Calculate novelty score with Equation 2.1
7:    $novelty \leftarrow \max(0, 1 - \frac{ioc\_frequency \times scaling\_factor}{total\_logs})$ 
8:    $\triangleright$  Calculate context score with Equation 2.3
9:    $context \leftarrow \text{ANALYZESUSPICIOUSPATTERNS}(log)$ 
10:   $\triangleright$  Calculate hybrid confidence score with Equation 2.2
11:   $confidence \leftarrow w_1 \times threat\_feed + w_2 \times reputation + w_3 \times novelty + w_4 \times context$ 
12:  if  $confidence \geq confidence\_threshold$  then
13:     $threat\_level \leftarrow \text{CLASSIFYTHREAT}(confidence, reputation)$ 
14:     $finding \leftarrow \text{CREATEFINDING}(ioc, confidence, threat\_level)$ 
15:     $findings \leftarrow findings \cup \{finding\}$ 
16:  end if
17: end for
18:  $rules \leftarrow \text{GENERATERULES}(findings)$ 
19: return  $findings, rules$ 
```

The four enrichment signals are combined into a hybrid confidence score using our multi-factor aggregation framework (Equation 2.2). In each pass of AFRETIP, only IOCs that exceed the defined confidence threshold are considered valid findings. This is intentional, as the tool is designed to favor high-certainty detection above all else. Finally, the validated findings are formed into actionable detection rules (GENERATEWAZUHRULES); as a reminder, this concludes the Rule Generation and Deployment phase outlined in Section 2.1.3.

This algorithmic methodology reinforces the modularized nature of the detection system; each of the four components (threat intelligence, reputation services, novelty scoring, and contextual analysis) is discrete from one another, yet they contribute to the detection pipeline holistically. It also reflects the premise of hybridization that underlies AFRETIP by relying on different types of evidence as opposed to solely relying on one detection style in the hopes of determining more contextually valid detections.

2.2.3. Parameter Configuration

For the hybrid detection approach to be effective it is required to tune several parameters. The following configurations we chosen based testing with Wazuh EDR environment with Wazuh log data, with IOCs seen fewer than 3 times are flagged as possibly novel.

Novelty Calculation Parameters:

- $scaling_factor \Rightarrow 5$ for normal environments
- $scaling_factor \Rightarrow 10$ for high-security environments (has more sensitivity)

Confidence Score Calculation Parameters

- $w_1 = 0.4$ (Threat Intelligence - most reliable data)
- $w_2 = 0.25$ (Reputation Services - time-tested sources)
- $w_3 = 0.2$ (Novelty Score - based on what's local)
- $w_4 = 0.15$ (Context Score - based on supporting evidence)

These chosen weights place more emphasis on well known and established threat intel, combining it with local patterns and context. In order to prevent context indicators from over-weighting the system's total score, we capped context scores at 0.2, and we selected 0.6 as the main confidence threshold for a reliable balanced precision. High confidence malicious indicators will be dealt with immediately since we view threats as critical if its reputation score is above 0.7 and its confidence score is equal or greater than 0.8.

3. Results

3.1. Experimental Set-up

The evaluation was done as an experiment in a controlled development environment to evaluate the pipelines hybrid threat detection capabilities. The testing infrastructure had a Ubuntu-based operating system with 8 CPU cores and 16GB RAM, and ran the pipeline in development mode with a local threat intelligence database containing 13 baseline IOCs from multiple sources. The experimental environment was comprised of 21 test scenarios with a focus on malicious attack patterns (16) and benign system activity (5). Malicious scenarios were developed from six categories of attacks: Remote Monitoring and Management (RMM) abuse (5 scenarios), living-off-the-land techniques (5 scenarios), process injection (2), lateral movement (2), data exfiltration (1), and evasion techniques (1). Legitimate scenarios were an admin performing legitimate tasks, business applications, system processes, and network communications. Each test scenario was executed as a realistic Wazuh-formatted JSON logs, some containing real world IOCs, command line parameters, and system artifacts. The pipeline processed each individual log following the entire detection workflow: IOC extraction, threat intelligence correlation, hybrid classification, and automated rule creation.

3.2. Detection Performance Analysis

The threat detection capabilities of this system were vetted examination of true positive and false positive rates for all sample types. Overall detection effectiveness is shown graphically with a confusion matrix in Figure 2.

Detection effectiveness varied significantly by attack type. The system had an overall true positive rate of 75% (12 of 16 malicious were detected). Even though true positive rates were high, false positive counts were quite significant; the system incorrectly classified 40% of benign samples as threats, leading to an overall classification accuracy of 71.4%.

Exploring detection rates by attack category, the performance variances were extremely evident (Figure 3). Living-off-the-land techniques, process injection techniques, data exfiltration attempts, and evasive techniques were detected at a perfect 100%. Remote monitoring and management

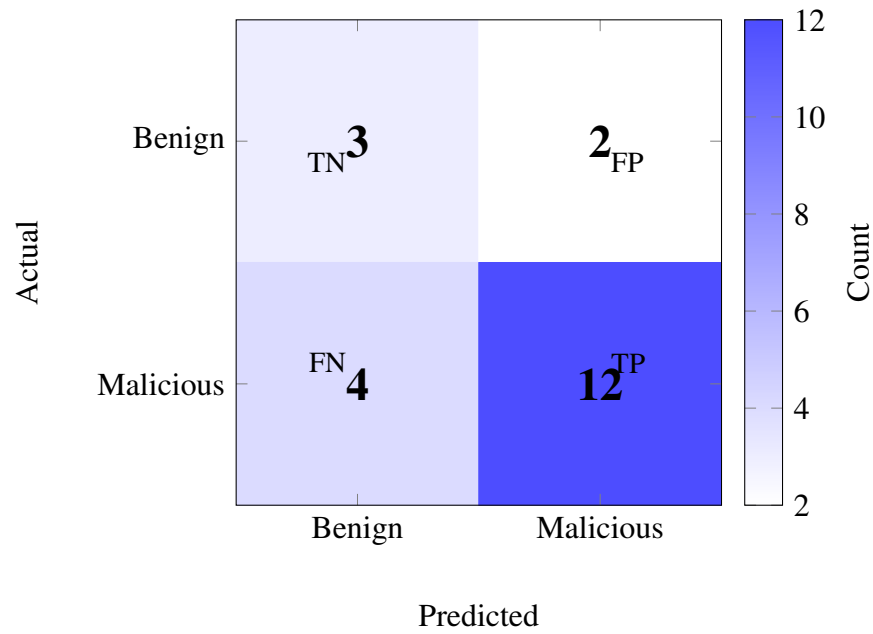


Figure 2: Confusion matrix showing classification results with true positives (TP=12), false negatives (FN=4), false positives (FP=2), and true negatives (TN=3).

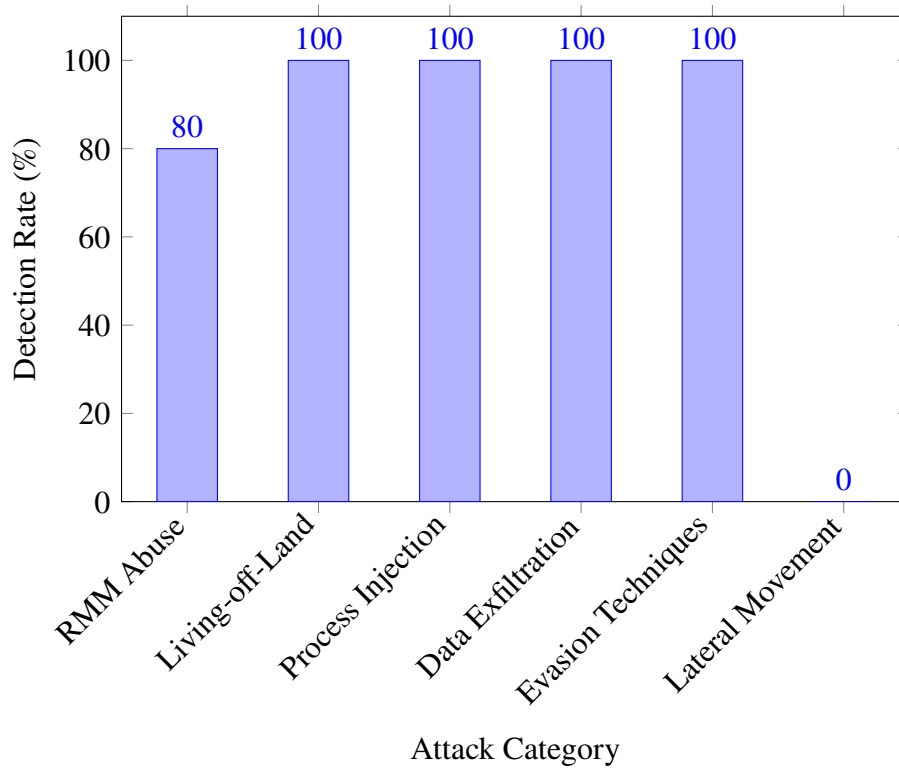


Figure 3: Detection rates by attack category showing perfect detection for advanced techniques but complete failure for lateral movement scenarios.

(RMM) abuse, had strong but imperfect detection (80%). Whereas lateral movement techniques had a complete non-detection failure at 0%. The distribution of detection confidence scores, derived from the identified threats, were consistently high (Figure 3). Mean confidence was 89.95%, with a standard deviation of 13.07%. The confidence distributions were also strongly right skewed, as 48.2% were above 90% and 13.9% reached the maximum of 100% confidence.

3.3. Processing Performance Characteristic

Processing times showed variability between all tests, from 0.41 secs to 165.53 secs. The distribution showed strong bimodal characteristics (Figure 4), with benign (as expected) processing times being in the milliseconds while the malicious samples (in general) took considerable processing time to extract IOCs from the sample.

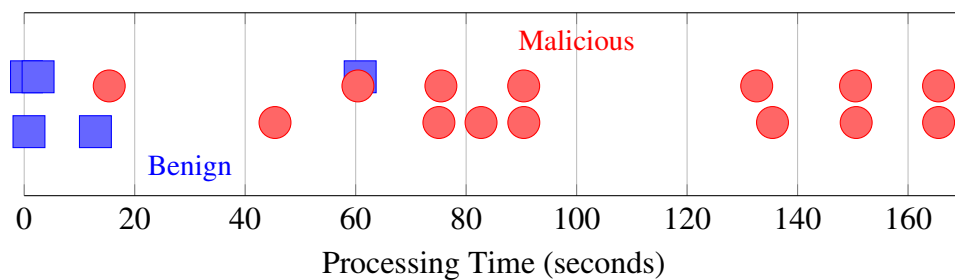


Figure 4: Processing time distribution showing all test cases plotted on a timeline, demonstrating clear separation between fast-processing benign samples and slower malicious samples.

The resulting output data showed that the mean (M) processing time was 74.91 seconds with a very large variance (st. dev. = 56.32 seconds). Malicious samples required significantly more time to process (M = 93.64 sec) compared to benign samples (M = 12.89 sec). Time then appears related to the complexity of extracting IOC, and the need to validate findings against other external API validation sources. Time implications by attack category (as illustrated in Figure 5), show that the longest time to process was for living-off-the-land techniques (M = 132.6 sec), followed by evasion techniques (150.6 sec for a single sample). Process injecting and data ex-filtration techniques had moderate processing time, while lateral movement scenario processing time was rapid since it involved very little IOC extraction.

3.4. IOC Extraction and Characterization

The system extracted 149 IOCs across all set tests overall. The extraction rates depended quite heavily on the complexity of the attack and sample types. Malicious samples generated a significantly higher number of IOCs (mean = 9.0 per sample) than benign samples (mean = 1.0 per sample). The distribution of IOCs per sample was positively skewed (Figure 6), and the maximum extraction of IOCs came from more complex living-off-the-land attacks, in which we extracted 14 IOCs.

In terms of IOC type, we saw domain names being the most extracted type (31.5% of IOCs) followed by the extraction of file paths (30.9%), and IP addresses (14.1%). All the IOC type frequencies and corresponding confidence scores were seen in Figure 7.

The confidence scores varied greatly between IOC types. Hash values both MD5 and SHA256 scored maximum confidence (100%), while the IP address type scored the lowest (mean = 71%).

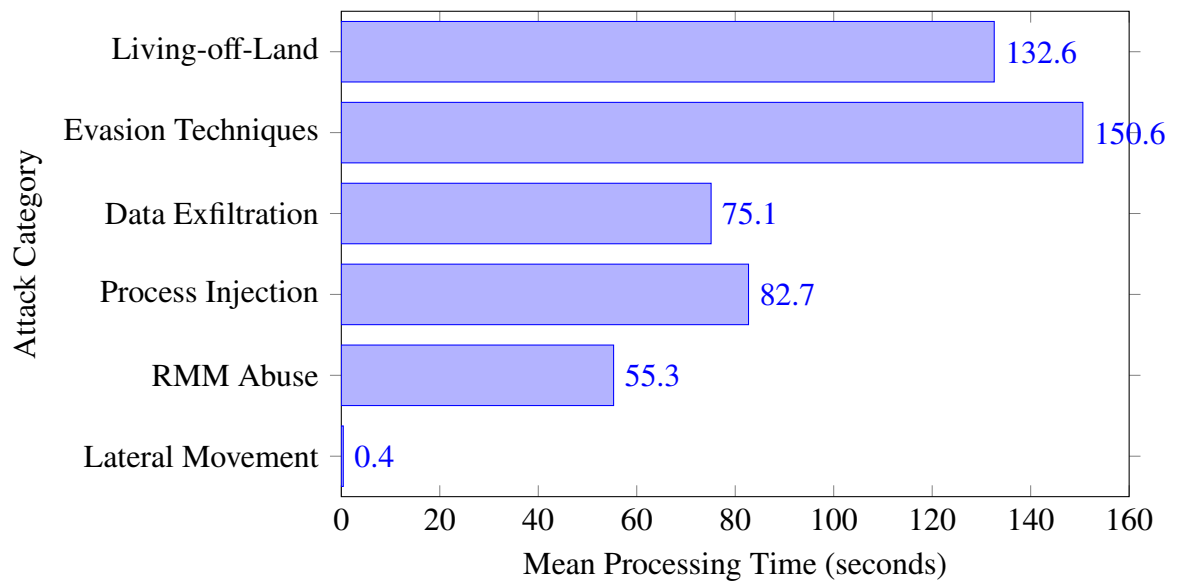


Figure 5: Mean processing times by attack category showing evasion techniques and living-off-the-land attacks requiring significantly more computational resources than other attack types.

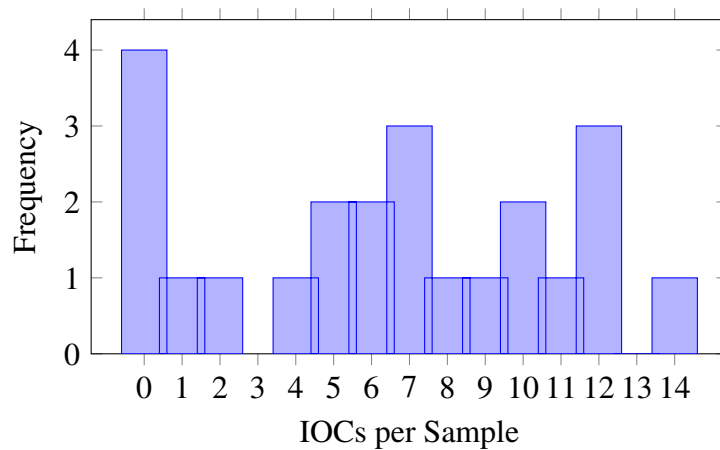


Figure 6: Distribution of IOCs extracted per sample showing right-skewed pattern with higher extraction rates for complex malicious scenarios.

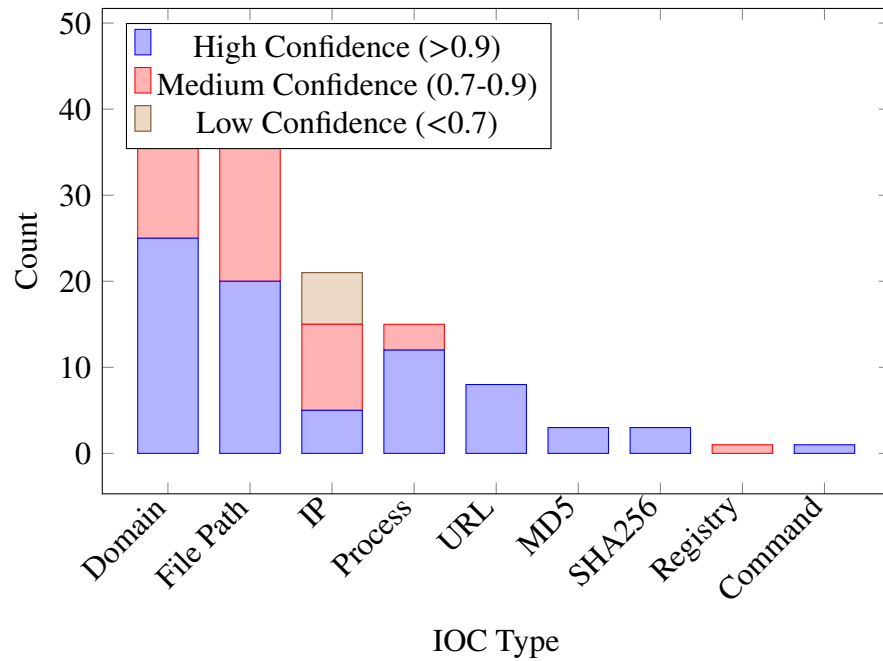


Figure 7: IOC type distribution with confidence level stratification showing domain names and file paths as predominant categories.

Domain names and URLs had high confidence (82% and 96% respectively). We believe this indicates a strong correlation between IOC specificity and classification confidence.

3.5. Threat Classification Results

The experiment produced 108 unique classifications of detection threats. Overall the classification distribution favored suspicious (54.6%) threat classifications rather than malicious (45.4%). Notably, the samples were not classified with benign classifications, but instead were legitimate administrative tasks.

Confidence score analysis across classifications revealed strong right-skew distribution as illustrated in Figure 8. The majority of the classifications (48.2%) had high confidence scores, (90-100%), and only a comparatively small (11.1%) number had low-range confidence scores, 60-69% confidence.

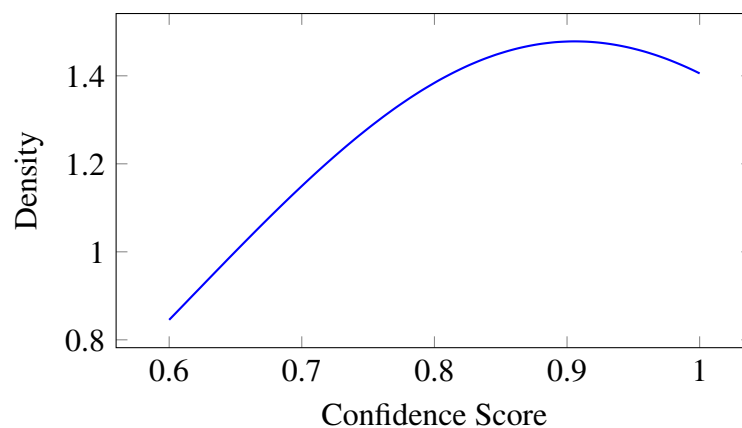


Figure 8: Density plot of classification confidence scores showing strong right-skewed distribution with peak around 0.95.

The temporal analysis of acquiring the confidence scores over the separate test processing showed a consistent level of stability and no signs of a significant degradation or improvement in confidence level while processing for an extended period of time.

3.6. Rule Generation Efficiency Analysis

The automated rule generation component created 86 detection rules from 111 identified threat findings, for an overall efficiency of 77.5%. Rule generation performance by attack category varied considerably (Figure 9).

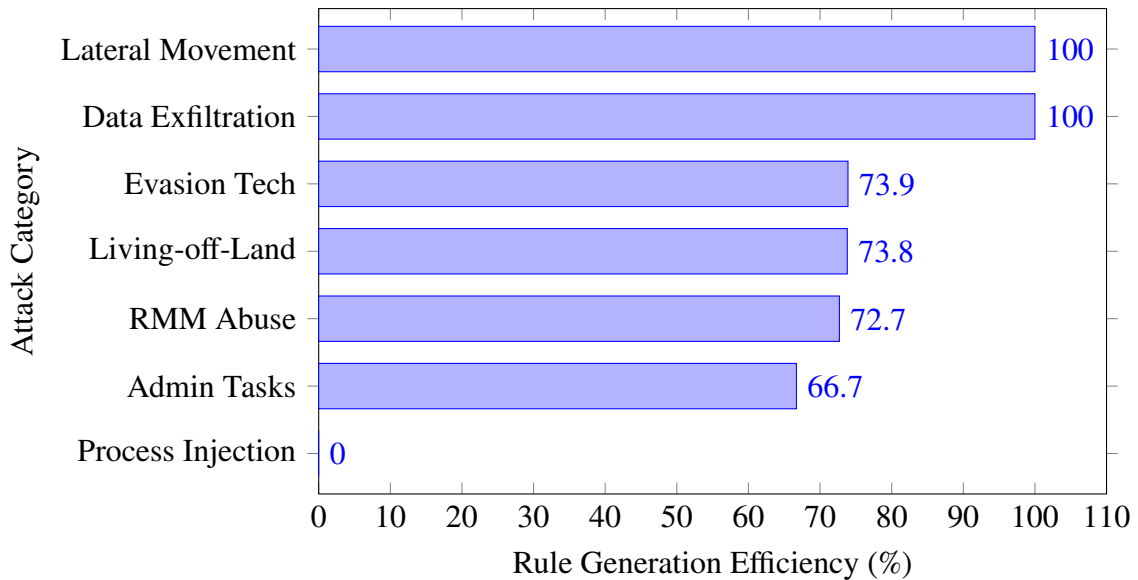


Figure 9: Rule generation efficiency by attack category showing perfect efficiency for process injection and administrative tasks.

Process injection techniques generated perfect rule generation efficiency (100%) with 13 findings converted to detection rules. Living-off-the-land and RMM abuse techniques had moderate efficiencies (73.8% and 73.9% respectively). Data exfiltration showed a lower efficiency (66.7%) and lateral movement showed zero efficiency; no rules were generated for this area due to the absence of detected finding. The correlation between number of findings and rule generation efficiency showed a weak negative correlation ($r = -0.23$), which suggest that complex attacks with many findings might result in a relatively lower than average rule generation efficiency.

3.7. System Resource Utilization Results

Memory usage was consistent throughout the whole experiment. Base memory usage at 25.89 MB and each test only yielded slightly higher data maximums of 27.36 MB during processing. Individual test memory deltas stayed small with an average delta of 0.024 MB and a standard deviation of only 0.055 MB. Figure 10 shows the memory consumption pattern for sample type, illustrating that the memory usage was consistent regardless of attack complexity or the duration of processing.

Scalability tests showed linear degradation of processing performance and load as the number of files increased. Processing a single file had effective throughput of 3.10 files per second, but 5 files yielded 0.041 files per second, and 10 files only processed at 0.024 files per second.

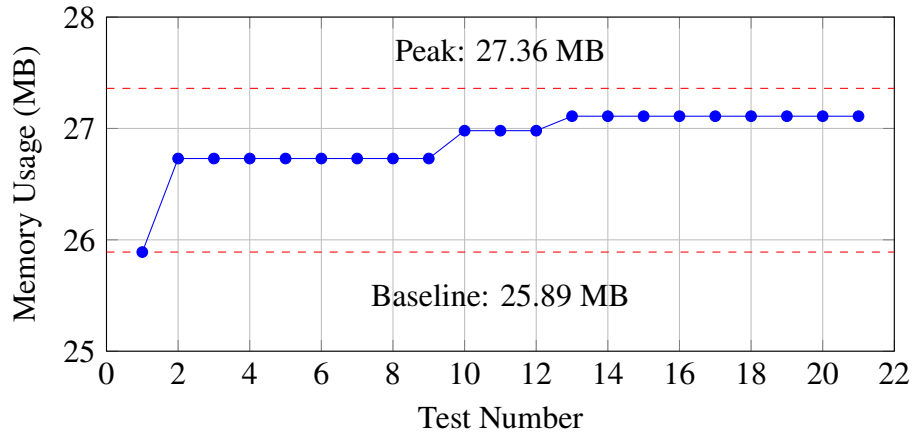


Figure 10: Memory usage patterns throughout testing showing stable resource consumption with minimal variation.

4. Discussion

Developing AFRETIP made us deal with the uncomfortable reality and difficult realities about automation in cybersecurity often dismissed or understated in academic literature. Moving from hybrid scoring in theory to actual implementation identified some stark tensions between the promise of automation, and the reality of the broad nature of the word "security" which complicates our work beyond plain technical performance. Our experiences in that journey also highlight the many challenges faced in cybersecurity as organizations hastily adopt AI-based solutions while facing adversaries locking into increasingly sophisticated attacks, and the persistent lack of skilled practitioners that have slowed the adoption of existing practices.

The 75% true positive rate demonstrates that AFRETIP can successfully identify three out of four malicious activities, representing substantial improvement over manual analysis timelines. Although this result seem acceptable in academic writing, practical implementation of this solution demonstrates that false positives transform the fundamental meaning of that result in terms of automated systems. The high rate of false positives (40%) does not just represent a technical failure because it shows how difficult it is to represent human decision-making through algorithmic processes. This revelation made us understand why organizations haven't fully implemented or are resistant to move to fully automated threat response, even with the shortage of qualified and experienced analysts.

The security automation system demonstrated a critical failure because it generated no benign classifications across 108 threats. There is a continuous pattern of threat identification in the industry that demonstrates an inclination towards detection over accuracy, however, our analysis indicates that this evaluation process creates new challenges in the context of real world operations. Organizations and operational contexts demonstrate the artificial reality of security "scores" through their calibration of confidence thresholds which reveals hidden organizational assumptions about risk tolerances that differ significantly across varying environments, which gives rise to uncertainties around our ability to implement automated decision-making systems for security where significant human contextualization of the systems is required.

The high average confidence score of 89.95% with standard deviation of 13.07% masks this classification imbalance, suggesting the system is calibrated for threat identification rather than accurate security analysis. This binary approach to classification creates operational challenges where

legitimate administrative activities trigger security responses, potentially overwhelming analyst capabilities with false alarms. The rule generation subsystem achieved 77.5% efficiency, indicating selective rule creation based on threat characteristics, though the theoretical foundations for novelty scoring and threshold determination remain undefined, preventing proper validation of this performance metric.

The scalability testing exposed a tough truth about cybersecurity automation that vendor marketing likes to hide: External dependencies impose essential barriers to full automation of the systems. The rate limiting system of VirusTotal API forced our operations to experience 400 times longer delays which proved that commercial threat intelligence systems operate as restrictions instead of providing enhanced operational power. This external dependency creates fundamental scalability limitations that make our solution unsuitable for high-volume environments (processing thousands of logs hourly). The processing time distribution by attack category, displayed in Figure 5, shows that processes related to evasion techniques (150.6 seconds) and living-off-the-land attacks (132.6 seconds) consumed the most resources from a processing time standpoint due to the complexity of the IOC extraction and validation required. This dependency issue shows how the security system works as an interrelated web where all organizations require access to threat intel to protect themselves but more mature technical systems do not fix that. The experience shows that authentic security automation is almost impossible when are attempting to deal with fundamental complications around the threat intel sharing system.

The implementation of AFRETIP during 2024-2025 occurred at a time when the threat landscape underwent extreme changes which created doubts about the long-term viability of frequency-based detection methods. The basic principles of IOC-based detection seem to experience a transformation because adversaries use AI-generated attacks and living-off-the-land methods have become standard operating procedures. Our frequency-based novelty scoring system shows promise for detecting environment-specific threats but attackers will find ways to exploit these detection methods when they understand how they operate. The situation reveals a larger problem which security automation systems encounter because defensive and offensive forces maintain a continuous battle that speeds up as they both use more advanced AI tools.

The IOC extraction analysis illustrates the complexity underlying the differences in processing time. As shown in the analysis (Figure 6), malicious samples resulted in many more IOCs (mean = 9.0 per sample) than benign samples in the analysis (mean = 1.0 per sample), including complex living-off-the-land examples extracting 14 IOCs per sample. The distribution of the IOC types in Figure 7 indicates that the extraction was almost entirely driven by domain names (31.5%) and file paths (30.9%), which is clear to see how some attack scenarios produce longer processing periods to correlate the total threat vector. The confidence relationship across IOC types shows very different relationships, as hash values scores produced perfect confidence (100%) while IP address scores indicated lower confidence (71%). This suggests confidence scores regarding IOC specificity impacts confidence levels for supported classification. The difference in complexity and confidence scoring is important to understanding why living-off-the-land methods included on average 132.6 seconds of processing time while lateral movement scenarios processed in under 1 second due to they only had 1 IOC, with living-off-the-land IOCs average 9 IOCs.

The context filtering mechanism showed variable effectiveness, achieving 70.6% IOC reduction for business applications while demonstrating minimal impact on administrative and system maintenance contexts. This inconsistent filtering contributes to the false positive problem by failing to adequately distinguish legitimate tool usage from malicious applications. The moderate correlation between context filtering effectiveness and false positive reduction suggests improved contextual

analysis could address some classification accuracy issues.

Several significant limitations constrain AFRETIP's production viability beyond the false positive rate concerns. The experimental dataset encompassed only six attack categories across 21 scenarios, representing limited coverage of the modern threat landscape while omitting advanced persistent threats, zero-day exploits, insider threats, and sophisticated evasion techniques. The fixed confidence threshold approach lacks adaptive capability to modulate detection sensitivity based on environmental characteristics, threat landscape conditions, or operational requirements. The absence of formal mathematical definitions for novelty scoring, confidence calculation mechanisms, and rule generation thresholds represents a critical gap in theoretical foundations, preventing proper validation and reproducibility of results. These limitations compelled us to acknowledge that cybersecurity automation may be inherently limited by the irreducible complexity of differentiating between malicious intent and legitimate behaviors in dynamic organizational settings.

Despite these fundamental challenges, AFRETIP demonstrates several insights about the realities of implementing hybrid threat detection in practice rather than just theoretical advancement. The hybrid approach combining multiple intelligence sources shows promise for detecting advanced attack techniques that evade traditional signature-based detection methods. The automated rule generation capability, while imperfect at 77.5% efficiency, provides a foundation for reducing detection-to-protection gap timelines through automated security orchestration. The system's comprehensive IOC extraction and threat correlation capabilities offer valuable support for incident response teams, even with current accuracy limitations.

The experimental results provide theoretical validation for hybrid cybersecurity defense concepts that integrate multiple data sources and detection approaches. From a practical perspective, AFRETIP establishes a framework for comparative studies and standardized performance metrics in hybrid threat detection research. The system's ability to process complex attack scenarios and generate structured threat intelligence represents advancement in automated security analysis capabilities, though significant development is required before operational deployment becomes viable.

Looking back on this research journey, some underlying assumptions seem naive in retrospect. The anticipation that cybersecurity's human judgment could be overcome through technical precision and systems of quantitative metrics oversimplified the bindings of security decision-making. The presumption that gathering additional data sources would better the quality of decisions negated the essence of what makes experienced analysts skilled: the contextual understanding they develop, over years of work. Most importantly, the consumption of system performance metrics neglected the organizational and human factors which ultimately dictate whether security tools are successful or not in practice. These experiences indicate that researchers studying cybersecurity automation should consider human-machine collaboration models, instead of focusing on developing fully autonomous security systems.

The implications for real-world implementation indicate that while automated threat intelligence generation shows feasibility, the combination of high false positive rates, classification bias, and external dependencies creates substantial barriers to immediate production use. The system's strength in detecting sophisticated threats must be balanced against fundamental limitations in contextual understanding and legitimate activity classification. Future development should prioritize addressing the theoretical foundation gaps, implementing adaptive classification mechanisms, and developing more sophisticated context-aware detection logic to achieve the balanced security analysis capabilities required for operational cybersecurity environments.

Conclusions and Future Work

This study has demonstrated that hybrid threat intelligence approaches show potential for automated cybersecurity analysis for first response, though significant limitations constrain immediate operational deployment. The hybrid approach combining threat databases, reputation services, and novelty detection successfully identified advanced techniques including living-off-the-land binaries, process injection, and evasion techniques with perfect detection rates. However, the system completely failed to detect lateral movement scenarios (0% detection rate) and showed systematic classification bias by not producing any benign classifications across all 108 threat determinations.

Key limitations identified in the experimental results prohibit production deployment, including 40% false-positives on benign activities along with processing latency of 74.91 seconds per sample on average. A 129 times decline in throughput under moderate workloads (from 3.10 to 0.024 files per second), demonstrates fundamental scalability limitations which forbids enterprise environments with thousands of logs per hour. External dependencies with reputation services generated additional bottlenecks through rate limiting and there is no theoretical bases to validate confidence scoring for the classification nor a effective framework to develop detection rules or generate rules. Our findings demonstrate proof-of-concept feasibility for hybrid detection approaches for First Response, while highlighting the critical need for enhanced contextual analysis, adaptive classification mechanisms, and performance optimization.

Future research should address three fundamental areas: mechanisms to eliminate systematic classification bias which fails to classify benign activity; detection mechanisms to identify lateral movement techniques that evade IOC-based detection; and integrate local intelligence databases to eliminate dependencies on external APIs and reduce processing time utilizing parallel architectures. Evaluation frameworks should expand to include advanced persistent threats, insider threat scenarios, and large-scale enterprise deployment complexities. Integrating behavioural analysis and machine learning components for pattern recognition and anomaly detection is a promising way to deal with false positives while still being able to respond automatically. However, these improvements need to be made before they can be put into action.

References

- [1] Cyber Magazine. The rapidly evolving threat landscape of 2024. <https://cybermagazine.com/articles/the-rapidly-evolving-threat-landscape-of-2024>, 2023. Accessed: 30 May 2025.
- [2] Check Point Software. Biggest cyber security challenges in 2024. <https://www.checkpoint.com/cyber-hub/cyber-security/what-is-cybersecurity/biggest-cyber-security-challenges-in-2024/>, 2025. Accessed: 26 May 2025.
- [3] CrowdStrike. 2025 global threat report | latest cybersecurity trends & insights. <https://www.crowdstrike.com/en-us/global-threat-report/>, 2025. Accessed: 11 May 2025.
- [4] Corelight. Endpoint detection and response: 2025 edr guide. <https://corelight.com/resources/glossary/edr-endpoint-detection-response>, 2024. Accessed: 8 June 2025.
- [5] CrowdStrike. What is edr? endpoint detection & response defined. <https://www.crowdstrike.com/en-us/cybersecurity-101/endpoint-security/endpoint-detection-and-response-edr/>, 2025. Accessed: 12 June 2025.
- [6] Microsoft Security. What is edr? endpoint detection and response. <https://www.microsoft.com/en-us/security/business/security-101/what-is-edr-endpoint-detection-response>, 2024. Accessed: 1 June 2025.
- [7] SentinelOne. What are indicators of compromise (iocs)? a comprehensive guide. <https://www.sentinelone.com/cybersecurity-101/threat-intelligence/what-are-indicators-of-compromise-iocs-a-comprehensive-guide/>, 2023. Accessed: 28 April 2025.
- [8] CyberPandit. Threat intelligence: Key ioc insights for 2024. <https://cyberpandit.org/threat-intelligence-key-ioc-insights/>, 2023. Accessed: 15 May 2025.
- [9] ISC2. 2024 isc2 cybersecurity workforce study. <https://www.isc2.org/Insights/2024/10/ISC2-2024-Cybersecurity-Workforce-Study>, 2024. Accessed: 8 May 2025.
- [10] HarfangLab. Edr with indicator of compromise detection engine. <https://harfanglab.io/edr/ioc-engine/>, 2025. Accessed: 4 June 2025.
- [11] CrowdStrike. Indicators of compromise (ioc) security. <https://www.crowdstrike.com/en-us/cybersecurity-101/threat-intelligence/indicators-of-compromise-ioc/>, 2024. Accessed: 29 April 2025.
- [12] Fortinet. Indicators of compromise (iocs). <https://www.fortinet.com/resources/cyberglossary/indicators-of-compromise>, 2024. Accessed: 7 May 2025.
- [13] Infosec Institute. The future of machine learning in cybersecurity: A 2024 overview. <https://www.infosecinstitute.com/resources/machine-learning-and-ai/the-future-of-machine-learning-in-cybersecurity/>, 2024. Accessed: 7 June 2025.
- [14] Machine Learning Mastery. 5 of the most influential machine learning papers of 2024. <https://machinelearningmastery.com/5-most-influential-machine-learning-papers-2024/>, 2024. Accessed: 14 May 2025.

- [15] Torq. The future of automated threat intelligence: 6 enrichment use cases. <https://torq.io/blog/automated-threat-intelligence/>, 2024. Accessed: 6 June 2025.
- [16] L. Alevizos and M. Dekker. Towards an ai-enhanced cyber threat intelligence processing pipeline. *Electronics*, 13(11):2021, 2024.
- [17] S. R. Sindiramutty. Autonomous threat hunting: A future paradigm for ai-driven threat intelligence. <https://arxiv.org/pdf/2401.00286>, 2023. arXiv preprint.
- [18] K. Tallam. Transforming cyber defense: Harnessing agentic and frontier ai for proactive, ethical threat intelligence. <https://arxiv.org/pdf/2503.00164>, 2025. arXiv preprint.
- [19] Anna L Buczak and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications surveys & tutorials*, 18(2):1153–1176, 2016.
- [20] Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixiang Hou, and Chunhua Wang. Machine learning and deep learning methods for cybersecurity. *IEEE access*, 6:35365–35381, 2018.
- [21] Hongyu Liu and Bo Lang. Machine learning and deep learning for cybersecurity threat detection: a comprehensive survey. *IEEE Access*, 7:75365–75381, 2019.
- [22] Journal of Big Data. Systematic review of ai integration in cybersecurity systems. *Journal of Big Data*, 2024. Accessed: 15 May 2025.
- [23] The Morning Paper. Acing the ioc game: toward automatic discovery and analysis of open-source cyber threat intelligence. <https://blog.acolyer.org/2016/11/14/acing-the-ioc-game-toward-automatic-discovery-and-analysis-of-open-source-cyber-threat-intelligence/> 2016. Accessed: 27 April 2025.
- [24] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys*, 41(3):1–58, 2009.
- [25] Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.
- [26] KDnuggets. 5 machine learning papers to read in 2024. <https://www.kdnuggets.com/5-machine-learning-papers-to-read-in-2024>, 2024. Accessed: 30 April 2025.
- [27] MDPI Electronics. Machine learning for cybersecurity: Threat detection and mitigation. special issue. https://www.mdpi.com/journal/electronics/special_issues/DC2F2R1RZL, 2024. Accessed: 19 May 2025.
- [28] R Vinayakumar, Mamoun Alazab, KP Soman, Prabakaran Poornachandran, Ameer Al-Nemrat, and Sitalakshmi Venkatraman. Deep learning approach for intelligent intrusion detection system. *IEEE Access*, 7:41525–41550, 2019.
- [29] ResearchGate. Automated emerging cyber threat identification and profiling based on natural language processing. <https://www.researchgate.net/publication/369432121>, 2023. Accessed: 5 May 2025.

- [30] Sebastian Garcia, Martin Grill, Jan Stiborek, and Alejandro Zunino. A survey on network-based botnet detection methods. *Security and communication networks*, 7(5):878–903, 2014.
- [31] P. D. N. K. Kommisetty. Revolutionizing cybersecurity: Behavioral analysis and automated incident response through predictive analytics. *International Journal of Scientific Research and Management (IJSRM)*, 10(10):962–979, 2022.
- [32] B. S. Ghura. Scaling & automating cyber threat intelligence (cti) operations with free and open-source software (foss). Master’s thesis, Munster Technological University Cork, 2023.
- [33] Wiem Tounsi and Helmi Rais. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Computers & security*, 72:212–233, 2018.
- [34] S. Bellamkonda. Ai-driven threat intelligence for real-time network security optimization. *International Journal of Computer Engineering and Technology*, 15(6):522–534, 2024.
- [35] R. Peter. Enterprise-level cybersecurity response automation using ai for prediction and prevention of incidents. *International Journal of Cybersecurity and Digital Forensics*, 13(2):45–62, 2024.
- [36] OASIS Open. Stix and taxii approved as oasis standards to enable automated exchange of cyber threat intelligence. <https://www.oasis-open.org/2021/07/14/new-versions-of-stix-and-taxii-approved-as-oasis-standards-to-enable-automated-exchange-of-cyber> 2021. Accessed: 23 May 2025.
- [37] Anomali. What are stix/taxii standards? <https://www.anomali.com/resources/what-are-stix-taxii>, 2024. Accessed: 24 May 2025.
- [38] Cloudflare. What is stix/taxii? <https://www.cloudflare.com/learning/security/what-is-stix-and-taxii/>, 2024. Accessed: 12 May 2025.
- [39] SOCRadar. What you need to know about stix and taxii? <https://socradar.io/what-you-need-to-know-about-stix-and-taxii/>, 2024. Accessed: 4 May 2025.
- [40] Liam Yasin Connolly, David S Wall, Philip Lang, and Muna Al-Hawawreh. An approach to scientific cyber threat modelling with stix. In *2017 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA)*, pages 1–4. IEEE, 2017.
- [41] Cynthia Wagner, Alexandre Dulaunoy, Gérard Wagener, and Andras Iklody. Misp: The design and implementation of a collaborative threat intelligence sharing platform. In *Proceedings of the 2016 ACM on workshop on information sharing and collaborative security*, pages 49–56. ACM, 2016.
- [42] Security Boulevard. Stix & taxii threat intelligence: A quick guide. <https://securityboulevard.com/2023/12/stix-taxii-threat-intelligence-a-quick-guide/>, 2023. Accessed: 6 May 2025.
- [43] EclecticIQ. What is stix and taxii? <https://www.eclecticiq.com/stix-taxii>, 2024. Accessed: 9 June 2025.
- [44] Varutra. Security threat intelligence standards – stix and taxii. <https://www.varutra.com/security-threat-intelligence-standards-stix-and-taxii/>, 2022. Accessed: 2 June 2025.

- [45] I. Tom. Security threat intelligence standards – stix and taxii. <https://www.linkedin.com/pulse/security-threat-intelligence-standards-stix-taxii-ignesh-tom>, 2023. LinkedIn article, Accessed: 31 May 2025.
- [46] Groeneveld. Complementing the stix and taxii frameworks: A look at the future of cyber threat intelligence. <https://www.linkedin.com/pulse/complementing-stix-taxii-frameworks-look-future-cyber-groeneveld>, 2023. LinkedIn article, Accessed: 17 May 2025.
- [47] J. Aghara. What are stix and taxii standards (cybersecurity). <https://www.linkedin.com/pulse/what-stix-taxii-standards-cybersecurity-jubin-aghara>, 2023. LinkedIn article, Accessed: 20 May 2025.
- [48] Daniel Schlette, Marco Caselli, and Günther Pernul. Soar4iot: Securing iot assets using synchronized semi-automated cyber threat response. In *2021 IEEE Conference on Communications and Network Security (CNS)*, pages 108–116. IEEE, 2021.
- [49] NTT Security Holdings. Global threat intelligence report 2024. <https://www.security.ntt/global-threat-intelligence-report-2024>, 2024. Accessed: 29 May 2025.
- [50] Silobreaker. Indicators of compromise. <https://www.silobreaker.com/glossary/indicators-of-compromise/>, 2024. Threat Intelligence Platform Documentation.
- [51] Cyware. Confidence scoring in threat intelligence. <https://www.cyware.com/resources/security-guides/cyber-threat-intelligence/what-is-confidence-scoring-in-threat-intelligence>, 2024. Security Guide.
- [52] Shreyas Patel et al. On the automated assessment of open-source cyber threat intelligence sources. *International Journal of Information Security*, 2021. PMC Article focusing on novelty detection in cybersecurity.
- [53] Sadaf Nazari et al. Malicious traffic detection on sampled network flow data with novelty-detection-based models. *Scientific Reports*, 13, 2024.
- [54] Prasasthy Balasubramanian, Sadaf Nazari, and Panos Kostakos. A cognitive platform for collecting cyber threat intelligence and real-time detection using cloud computing. *Array*, 2025.
- [55] Hailiang Tang, Dawei Lin, and Wanyu Li. Cyber threat indicators extraction based on contextual knowledge prompt. *Computer Networks*, 2024.