Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review article

Evolving techniques in cyber threat hunting: A systematic review

Arash Mahboubi [a,*], Khanh Luong [a], Hamed Aboutorab [a], Hang Thanh Bui [a], Geoff Jarrad [b], Mohammed Bahutair [b], Seyit Camtepe [b], Ganna Pogrebna [a], Ejaz Ahmed [b], Bazara Barry [c], Hannah Gately [c]

[a] *Charles Sturt University, Australia*
[b] *CSIRO's Data61, Australia*
[c] *New South Wales Department of Customer Service, Australia*

ARTICLE INFO

Keywords:
Threat hunting
Hypothesis
Machine learning
OpenAI voice engine
Cyber threat intelligence
Generative AI

ABSTRACT

In the rapidly changing cybersecurity landscape, threat hunting has become a critical proactive defense against sophisticated cyber threats. While traditional security measures are essential, their reactive nature often falls short in countering malicious actors' increasingly advanced tactics. This paper explores the crucial role of threat hunting, a systematic, analyst-driven process aimed at uncovering hidden threats lurking within an organization's digital infrastructure before they escalate into major incidents. Despite its importance, the cybersecurity community grapples with several challenges, including the lack of standardized methodologies, the need for specialized expertise, and the integration of cutting-edge technologies like artificial intelligence (AI) for predictive threat identification. To tackle these challenges, this survey paper offers a comprehensive overview of current threat hunting practices, emphasizing the integration of AI-driven models for proactive threat prediction. Our research explores critical questions regarding the effectiveness of various threat hunting processes and the incorporation of advanced techniques such as augmented methodologies and machine learning. Our approach involves a systematic review of existing practices, including frameworks from industry leaders like IBM and CrowdStrike. We also explore resources for intelligence ontologies and automation tools. The background section clarifies the distinction between threat hunting and anomaly detection, emphasizing systematic processes crucial for effective threat hunting. We formulate hypotheses based on hidden states and observations, examine the interplay between anomaly detection and threat hunting, and introduce iterative detection methodologies and playbooks for enhanced threat detection. Our review encompasses supervised and unsupervised machine learning approaches, reasoning techniques, graph-based and rule-based methods, as well as other innovative strategies. We identify key challenges in the field, including the scarcity of labeled data, imbalanced datasets, the need for integrating multiple data sources, the rapid evolution of adversarial techniques, and the limited availability of human expertise and data intelligence. The discussion highlights the transformative impact of artificial intelligence on both threat hunting and cybercrime, reinforcing the importance of robust hypothesis development. This paper contributes a detailed analysis of the current state and future directions of threat hunting, offering actionable insights for researchers and practitioners to enhance threat detection and mitigation strategies in the ever-evolving cybersecurity landscape.

Contents

[*] Corresponding author.

E-mail addresses: amahboubi@csu.edu.au (A. Mahboubi), kluong@csu.edu.au (K. Luong), haboutorab@csu.edu.au (H. Aboutorab), hbui@csu.edu.au (H.T. Bui), geoff.jarrad@data61.csiro.au (G. Jarrad), mohammed.bahutair@data61.csiro.au (M. Bahutair), seyit.camtepe@data61.csiro.au (S. Camtepe), gpogrebna@csu.edu.au (G. Pogrebna), ejaz.ahmed@data61.csiro.au (E. Ahmed), bazara.barry@cyber.nsw.gov.au (B. Barry), hannah.gately@cyber.nsw.gov.au (H. Gately).

https://doi.org/10.1016/j.jnca.2024.104004
Received 12 May 2024; Received in revised form 30 July 2024; Accepted 18 August 2024
Available online 23 August 2024

1. Introduction

In the evolving landscape of cybersecurity, emerging threats, and lateral movement tactics pose significant challenges. Lateral movement refers to the techniques attackers use to navigate through a network after gaining initial access, seeking to elevate privileges, or access sensitive data. These threats often remain undetected for extended periods, enabling adversaries to embed themselves deeply within an organization's infrastructure. Such threats can be remotely activated by attackers or triggered by seemingly benign user actions, exploiting vulnerabilities, and leveraging legitimate processes and applications to evade detection. The stealthy nature of certain tactics — e.g. a stealthy lateral movement characterized by a set of unique features, such as requiring no elevated privileges, creating no new connections, necessitating no additional authentication, and involving no process injection — renders them undetectable by state-of-the-art (SOTA) detection mechanisms (Niakanlahiji et al., 2020). These tactics allow malicious actors to orchestrate sophisticated attacks, posing a substantial risk to information security.

The 2024 Global Threat Report has identified over 230 active adversaries and reported the fastest eCrime breakout time recorded at just 2 min and 7 s in 2023. This underscores the pressing need for rapid detection and response capabilities to mitigate these swift intrusions. The report highlights a significant rise in covert cyber activities, with substantial increases in data theft, cloud breaches, and malware-free attacks. These trends indicate that adversaries are continuously evolving and adapting despite advancements in detection technologies (CrowdStrike, 2024). Additionally, the adversarial use of generative AI (Artificial Intelligence) has escalated concerns regarding the development of highly convincing social engineering campaigns and the creation of malicious software, tools, and resources for more potent attacks. Trends from 2023 indicate that AI was frequently employed for social engineering, with the technology's capabilities offering adversaries endless opportunities to enhance their sophistication.

To counteract these advanced threats, a multifaceted approach is imperative. Strategies to mitigate attacks that involve emerging threats and lateral movement include the deployment of advanced intrusion detection systems (IDS) that employ behavioral analysis and anomaly detection techniques (Hossain et al., 2017). Implementing strict access controls and segmenting networks can also limit the ability of an attacker to move laterally. Additionally, continuous monitoring and real-time analysis of network traffic and user behavior can help identify suspicious activities early. Employing endpoint detection and response (EDR) solutions and conducting regular vulnerability assessments and penetration testing are also critical components of a robust cybersecurity posture. These measures aim to reduce the attack surface and detect threats before they can cause significant damage (Hassan et al., 2020).

However, despite these proactive and comprehensive approaches, the dynamic nature of cybersecurity threats often outpaces the capabilities of even the most advanced defensive measures. The aforementioned strategies are essential but not sufficient in isolation to address the full spectrum of new and emerging threats. This insufficiency stems from several critical factors. Attackers continuously adapt and evolve their tactics, techniques, and procedures (TTPs) (Alsaheel et al., 2021), making it difficult for static defense mechanisms to keep pace. The complexity of modern IT environments, characterized by cloud services, Internet of Things (IoT) devices, and remote access, significantly expands the attack surface, offering multiple vectors for exploitation (Jurcut et al., 2020). The multidimensional nature of the data generated by these complex environments complicates the task of effectively analyzing and identifying subtle and intricate patterns indicative of a breach (Tang et al., 2022). Additionally, the lack of resources and skilled analysts typically limits the ability of organizations to analyze and interpret the vast amounts of data generated by security tools, leading to potential oversight of critical threats.

In response to these challenges, the discipline of threat hunting has emerged as a critical component in the fight against sophisticated cyber threats. In an era dominated by big data — with issues such as unstructured data, and the heterogeneity of data sources — traditional security measures often fall short (Tang et al., 2022). Threat hunting involves proactively searching for cyber threats and suspicious activities that evade existing security solutions, leveraging known indicators of compromise (IoCs) and analyzing patterns that may suggest malicious activities (Gao et al., 2021a). The complexity and volume of data, combined with the need for timely detection and response, necessitate the use of advanced analytical techniques and the exploitation of legitimate processes (Botacin et al., 2021; Mahboubi et al., 2021) and applications to uncover stealthy attacks. This proactive approach helps organizations stay one step ahead of attackers, minimizing the risk of significant breaches before they manifest into full-blown attacks.

Industries across the spectrum are increasingly adopting threat hunting practices, guided by frameworks and standards developed by cybersecurity organizations. Tools such as Splunk, RedLine, and Loki (Milajerdi et al., 2019a) have become indispensable in this endeavor, providing data analytics capabilities to sift through large amounts of data in search of anomalies and IoCs. However, the detection of lateral movement remains a formidable challenge for current security models. The utilization of legitimate system processes and applications by attackers to conduct their operations complicates the identification of malicious activities. Consequently, threats often remain undetected, exploiting the blind spots of traditional security tools. The complexity of modern IT environments, coupled with the sophistication of attack techniques (Alzaabi and Mehmood, 2024), demands innovative approaches to detect these stealthy maneuvers.

In response to these challenges, the concept of developing an ‘avatar’ through threat hunting has gained traction. This approach aims to create a comprehensive understanding of current methodologies and existing challenges through empirical and experimental analysis (Horta Neto and Fernandes Pereira dos Santos, 2020). A review of the literature underscores the need for a paradigm shift towards more adaptive and intelligence-driven security strategies (Schlette et al., 2021b; Li et al., 2022a). Issues such as the lack of labeled datasets for machine learning (ML) models (Bae et al., 2024; Mikhail et al., 2020), over-reliance on statistical analysis (Jadidi and Lu, 2021), and the need for contextual understanding of network behavior highlight the gaps in current approaches.

1.1. Research objective and questions

This systematic review paper aims to investigate the rapidly evolving domain of cybersecurity, with a specific focus on threat hunting as a proactive defense mechanism against advanced cyber threats. It seeks to explore the evolution of threat hunting practices, assess the current state of research on formal mathematical hypotheses in threat hunting, analyze existing approaches and techniques developed by the research community, and outline the primary challenges faced by practitioners in the field. We pose the following research questions:

1. **RQ1**: How has the evolution of technology and methodologies impacted the development and practices of threat hunting over time?
2. **RQ2**: What is the current status of research on the development and application of formal mathematical hypotheses in threat hunting? Is there potential for formulating a comprehensive and adaptable mathematical model to enhance threat hunting practices?
3. **RQ3**: What methodologies and strategies have been devised by the research community for effective threat hunting?
4. **RQ4**: What are the primary challenges faced by professionals in the field of threat hunting, and what are the state-of-the-art approaches to address these challenges?

By addressing these questions, this paper aims to provide insights into the evolution, current state, methodologies, and challenges of threat hunting within the cybersecurity landscape.

1.2. Methodology

To answer research questions **RQ1**–**RQ4**, we conducted a systematic literature review (SLR) (Lame, 2019) staged across several phases, as shown in Fig. 1. In Phase 1, papers with any of the key phrases “threat hunting”, “threat detection”, “security information management”, “adversary tactics, techniques or procedures” or “cyber threat hunting techniques” in the title were automatically selected. The search was conducted on several scientific platforms, including IEEE, ACM, Web of Science, Scopus, and the Google Scholar search engines. Initially, 1696 papers were retrieved for the SLR.

In Phase 2, these papers were filtered by examining the abstracts to identify those that contained any of the following key phrases: “cyber threat intelligence”; “security analytic/s”; “proactive cybersecurity”; “incident response”; “behavioral analysis in cybersecurity”; “intrusion detection system”; “machine learning in cybersecurity”; “threat actor tactics and techniques”; “security information and event management”; “endpoint detection and response”; “event log”; and “log anomaly”. The inclusive and exclusive criteria used are shown in Fig. 1. This filtering resulted in 287 selected papers.

In Phase 3, a scanning phase was conducted by reviewing each selected paper to determine its relevance to threat hunting or threat detection; however, survey papers, posters, and letters were excluded. As a result, 117 papers were selected. To address research questions **RQ1**–**RQ4**, we categorized the selected papers as follows: 12 papers which focus on threat hunting procedures were selected to address **RQ1**, as discussed in detail in Section 2. Section 3 clarifies the hypothesis formulation of 4 of the selected papers in their threat hunting models to address **RQ2**. In Section 4, we analyze the techniques used in 63 papers to answer **RQ3**. Finally, in Section 5, we identify the challenges in threat hunting techniques, tactics, and procedures through a review of 38 papers to answer **RQ4**.

The structure of the paper is as follows. In Section 2, we explain the background of threat hunting. In Section 3, we explain the hypothesis formulation. In Section 4, we review the existing threat hunting approaches in the literature. Section 5 analyzes the challenges of threat hunting models. In Section 6, we investigate the existing surveys with a focus on threat hunting. Finally, Section 7 discusses the findings, followed by Section 8, which concludes the paper.

2. Background

This section offers an in-depth exploration of the evolutionary trajectory of threat hunting, addressing research question RQ1. The evolution of threat hunting in the realm of cybersecurity illustrates a dynamic shift towards proactive defense strategies against increasingly sophisticated cyber threats. Originating from the broader field of cybersecurity and network defense, the concept of threat hunting does not have a single point of origin or creator. Instead, it represents a collective response to the need for more advanced methods of detecting and neutralizing cyber threats. Gaining formal recognition in the early 2010s, threat hunting embodies a proactive and cyclical process aimed at preemptively identifying and mitigating potential cyber threats in complex network environments, such as those found in enterprise settings. This approach involves the formulation, testing, and refinement of hypotheses about potential threats, leveraging practices such as deploying specialized monitoring tools in specific network segments to gather information and validate or revise initial hypotheses (Shu

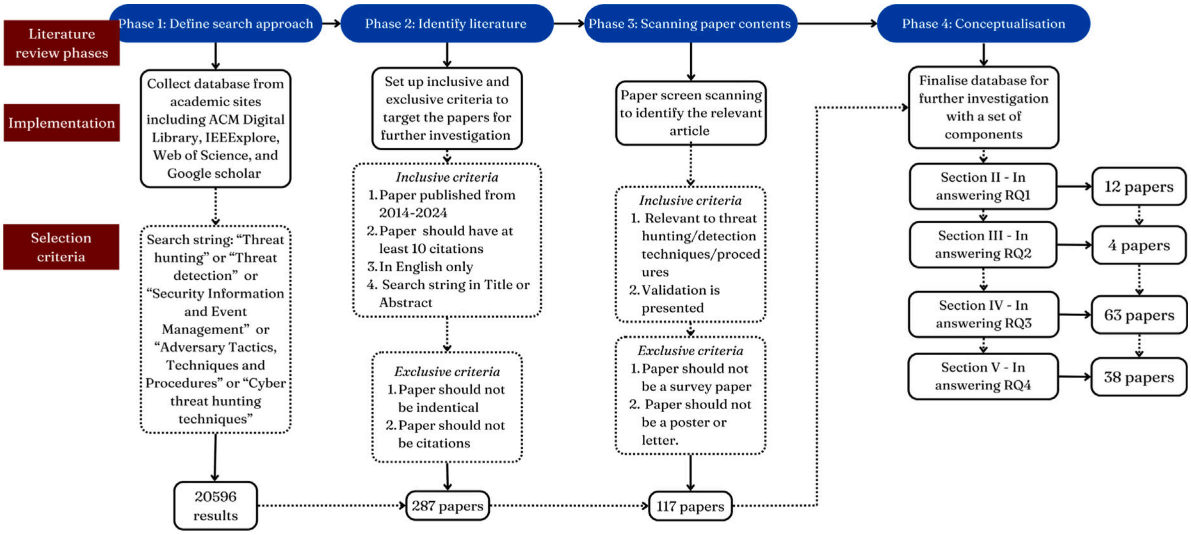
Literature review phases
Implementation
Selection criteria
Phase 1: Define search approach
Phase 2: Identify literature
Phase 3: Scanning paper contents
Phase 4: Conceptualisation
Collect database from academic sites including ACM Digital Library, IEEExplore, Web of Science, and Google scholar
Search string: "Threat hunting" or "Threat detection" or "Security Information and Event Management" or "Adversary Tactics, Techniques and Procedures" or "Cyber threat hunting techniques"
20596 results
Set up inclusive and exclusive criteria to target the papers for further investigation
Inclusive criteria
1. Paper published from 2014-2024
2. Paper should have at least 10 citations
3. In English only
4. Search string in Title or Abstract
Exclusive criteria
1. Paper should not be indentical
2. Paper should not be citations
287 papers
Paper screen scanning to identify the relevant article
Inclusive criteria
1. Relevant to threat hunting/detection techniques/procedures
2. Validation is presented
Exclusive criteria
1. Paper should not be a survey paper
2. Paper should not be a poster or letter.
117 papers
Finalise database for further investigation with a set of components
Section II - In answering RQ1
Section III - In answering RQ2
Section IV - In answering RQ3
Section V - In answering RQ4
12 papers
4 papers
63 papers
38 papers

Fig. 1. SLR methodology.

et al., 2018). This methodology seeks not only to uncover and intercept attacks early but also to identify unusual behaviors that indicate the presence of harmful entities, thereby enhancing the overall security posture of organizations (Saeed et al., 2023; Mansfield-Devine, 2017).

The incorporation of modern security solutions such as Security Information and Event Management (SIEM) (Martins and Medeiros, 2022), and Security Orchestration, Automation, and Response (SOAR) (Schlette et al., 2021a) into threat hunting practices underscores the importance of leveraging technology to improve detection capabilities and investigative efficiency. These systems offer functionalities that range from vulnerability tracking to event analysis and digital forensics. However, for these tools to effectively contribute to the enhancement of enterprise security defenses, they must be integrated with threat hunting capabilities that enable the proactive discovery of threats (Liu et al., 2018; Bowman et al., 2020).

One crucial aspect of threat hunting is the differentiation and understanding of threat indicators, which are pivotal in identifying potential security breaches. The Detection Maturity Level (DML) model (Mavroeidis and Bromander, 2017) emphasizes the variability in the semantic levels of threat indicators, with higher semantic indicators such as goals, strategies, or TTPs being more valuable than lower semantic indicators like network artifacts and atomic indicators, including IP addresses. This model highlights the necessity for SIEM tools to evolve beyond providing low-level indicators, advocating for capabilities that can offer insights into higher-level threat indicators (Bromander et al., 2016).

Threat indicators are broadly categorized into Indicators of Compromise (IoCs) (Jadidi and Lu, 2021). IoCs signal that an incident has already occurred, thereby placing the organization in a reactive stance. Examples of IoCs include unusual network traffic, suspicious user account activity, and login anomalies. Alternatively, Indicators of Concern, often derived from open-source intelligence (Johnsen and Franke, 2019), involve collecting data from publicly available sources to aid in cyberattack detection and threat hunting. This distinction underscores the varied approaches to identifying and responding to cyber threats, highlighting the critical role of intelligence and proactive measures in modern cybersecurity practices.

It is important to note that the threat hunting framework differs fundamentally from an incident-response process. The key distinction lies in the nature of their approaches: incident response is a reactive model that addresses security incidents after they have occurred; whereas threat hunting is proactive, aiming to identify and mitigate threats before they can inflict harm (Schlette et al., 2021a). Specifically, an incident-response strategy involves a sequence of actions executed by a security team following a security breach (e.g. an attack) with objectives centered around limiting the impact, reducing recovery time, and cutting down associated costs (IBM, 2024). This plan typically encompasses steps for detection, digital forensic analysis, investigation, and recuperation from security breaches. In contrast, threat hunting is an assertive security approach where security professionals actively examine system data, leveraging their expertise to spot threats that might have eluded standard security measures. This approach entails formulating and investigating hypotheses about potential attacks, grounded in an understanding of the system's architecture and the broader threat environment, with the goal of uncovering previously unidentified threats (Kaiser et al., 2023).

Threat hunting remains a relatively underdeveloped and vaguely delineated concept in terms of both its procedures and its integration within organizational structures. Currently, the predominant approach in many organizations is to respond to alerts and incidents reactively rather than adopting a proactive stance in identifying potential threats (Nour et al., 2023). Research highlights some compelling points, notably the interdependence of threat intelligence and threat hunting and its impact on effective performance, emphasizing the need for more automation in these processes. Furthermore, these studies indicate that organizations which incorporate tangible enhancements to their security procedures typically observe improvements in terms of speed and accuracy. Notably, the adoption of proactive threat hunting practices has been linked to reduced exposure to security threats, demonstrating its potential benefits.

2.1. Threat hunting versus anomaly detection

In the dynamic landscape of cybersecurity, the interplay between offensive maneuvers and defensive strategies necessitates a continual evolution of tactics and tools. Proactive measures in cybersecurity are essential for defending against potential threats. Two primary approaches to detect and mitigate cyber risks are those of *anomaly detection* and *threat hunting*. While both methods contribute to enhancing the security of the environment, they operate differently and serve distinct purposes.

Anomaly detection is a largely automated technique used to identify deviations from normal behavior within a system or network. It relies on statistical analysis, ML algorithms, and predefined thresholds to flag activities or events that differ significantly from expected patterns. This method can be effective for detecting subtle anomalies that may go unnoticed by manual inspection, such as insider threats or stealthy attacks that blend into legitimate system events. Anomaly detection can be both *reactive*, where it identifies deviations after they occur, and *proactive*, where it establishes baselines of normal behavior and alerts when deviations exceed predefined thresholds. Furthermore, anomaly detection systems can adapt dynamically to evolving threats by adjusting detection thresholds and models based on the changing characteristics of the environment.

The semi-automated aspect of anomaly detection systems makes them suitable for monitoring large volumes of data in real time. However, by this very fact, such systems may generate many false positives, as legitimate activities that deviate from established patterns can also trigger alerts. It is essential to tune thresholds and refine algorithms to minimize false positives.

Threat hunting is a *proactive* cybersecurity strategy aimed at identifying and mitigating potential threats that may have bypassed traditional security measures. It involves searching through networks to detect and isolate advanced threats that elude traditional defensive mechanisms such as IDS, IPS, and firewalls. This process involves extensive data analysis across various sources, including network traffic, system logs, endpoint telemetry, and threat intelligence feeds, to uncover malicious activities.

What distinguishes threat hunting is its reliance on the ***hypothesis-driven*** exploration of threats, leveraging human intelligence and intuition to anticipate and uncover sophisticated cyber attacks (Alevizos and Dekker, 2024). This approach not only contributes to a deeper understanding of adversarial tactics but also serves as a critical resource for the continuous improvement of automated defense systems.

Anomaly detection, statistical analysis, ML, and other cybersecurity technologies can all benefit significantly from the insights garnered through threat hunting activities (Rashid et al., 2022). These technologies, while powerful in their right, depend on high-quality, relevant data to effectively identify and mitigate threats. Threat hunters, by analyzing patterns of compromise within heterogeneous data sources (Samtani et al., 2020), often uncover IoCs and TTPs of adversaries that have successfully bypassed existing defensive measures. This intelligence is invaluable for updating the configurations and algorithms of anomaly detection systems and ML models, allowing them to adapt to the evolving threat landscape more adeptly.

2.2. A systematic threat hunting process

The SANS Institute's threat hunting maturity model categorizes organizations into five levels based on their threat hunting capabilities, namely:

Initial **(Level 0)** Reliance on automated reporting with little to no routine data collection.

Minimal **(Level 1)** Incorporation of threat intelligence indicator searches with moderate to high data collection.

Procedural **(Level 2)** Following analysis procedures created by others with high to very high data collection.

Innovative **(Level 3)** Creating new data analysis procedures with high to very high data collection.

Leading **(Level 4)** Automating successful data analysis procedures with high to very high data collection.

From these categories we have derived a systematic, ten-step process for threat hunting, as illustrated in Fig. 2 for an enterprise setting.

1. **Ingestion of Heterogeneous Data Sources:** This fundamental phase involves the collection of data from diverse inputs, such as data lakes, and system and network logs. This broad spectrum of data sources is crucial for capturing the complex and varied nature of cybersecurity threats, setting the stage for effective threat detection and classification.

2. **Formulation of Threat Definitions and Hypotheses:** Understanding what constitutes a threat within a specific operational context is critical. By generating hypotheses about potential threats, informed by current trends, intelligence, and system vulnerabilities, organizations can engage in targeted threat hunting, laying the groundwork for a nuanced classification of threats.

3. **Proactive Threat Hunting:** Armed with well-defined hypotheses, the proactive search for indicators that align with these hypotheses is undertaken. This stage is vital for intercepting threats before they materialize, ensuring that subsequent classification is both timely and relevant.

4. **Employment of Threat Observation Techniques:** Utilizing a variety of analytical tools and techniques, such as data analysis and anomaly detection, this phase facilitates the identification of anomalous activities indicative of cybersecurity threats, preparing the ground for their classification based on observed characteristics.

5. **Classification of Identified Threats:** Following identification, threats are categorized into clusters according to their shared characteristics, behaviors, or potential impacts. This classification not only enhances the efficiency of analysis and response strategies but also informs the refinement of threat definitions and hypotheses.

6. **Human Validation of Threats:** The role of human analysts is pivotal in validating the classification of threats, ensuring that the categorization accurately reflects the nature of the identified threats and minimizing false positives. This human insight is crucial for the reliable application of classification criteria.

7. **Assessment Against Prevention and Detection Mechanisms:** The validated and classified threats are assessed against existing security frameworks (e.g. IDS, IPS, and firewalls) to evaluate their effectiveness and to identify any security gaps. This assessment is informed by the detailed classification of threats, enabling targeted improvements.

8. **Extraction of Threat Signatures and Patterns:** Detailed analysis of classified threats allows for the extraction of distinctive features, signatures, and behavioral patterns, enriching the dataset of threat indicators and enhancing future threat detection efforts.

9. **Enhancement of Detection and Mitigation Frameworks:** Incorporating the latest threat intelligence, including APT tactics, MITRE ATT&CK techniques, and STIX data, into defense mechanisms ensures that the system's protection measures are current and effective. The classification of threats plays a critical role in identifying the specific system updates required.

10. **Iterative Enhancement Process:** Acknowledging the ever-evolving nature of the threat landscape, this approach emphasizes the importance of continually updating threat definitions, hypotheses, classification criteria, and mitigation strategies through an iterative process, ensuring that cybersecurity measures remain robust and responsive.

2.3. Augmented threat hunting methodologies

Cyber threat hunting methodologies have experienced significant evolution, transitioning from manual procedures to encompassing automated techniques (Nour et al., 2023). Initially, threat hunting predominantly constituted a manual endeavor, where a security analyst, leveraging their extensive knowledge and comprehension of the network architecture and behavior, engaged in the analysis of various data streams to construct hypotheses concerning potential cyber threats. Such manual investigations typically involved the identification of patterns indicative of lateral movements by threat actors, among other sophisticated attack vectors obtained from OSCTI platforms (Gao et al., 2021a).

To augment the process of hunting threats (e.g. AI, cyber, and computer crimes — see Fig. 3), there has been a shift towards the integration of automation and machine assistance for both attackers and defenders (Kaloudi and Li, 2020). This paradigm shift allows defenders to leverage advanced algorithms and ML models to sift through vast quantities of data at speeds and volumes unattainable by human analysts alone. The automated systems are designed to detect anomalies, patterns, and IoCs that might suggest the presence of a cybersecurity threat. However, attackers may also use AI technology-enhanced learning methods, turning them into tools for automated attacks. The transition towards AI technologies that have the ability

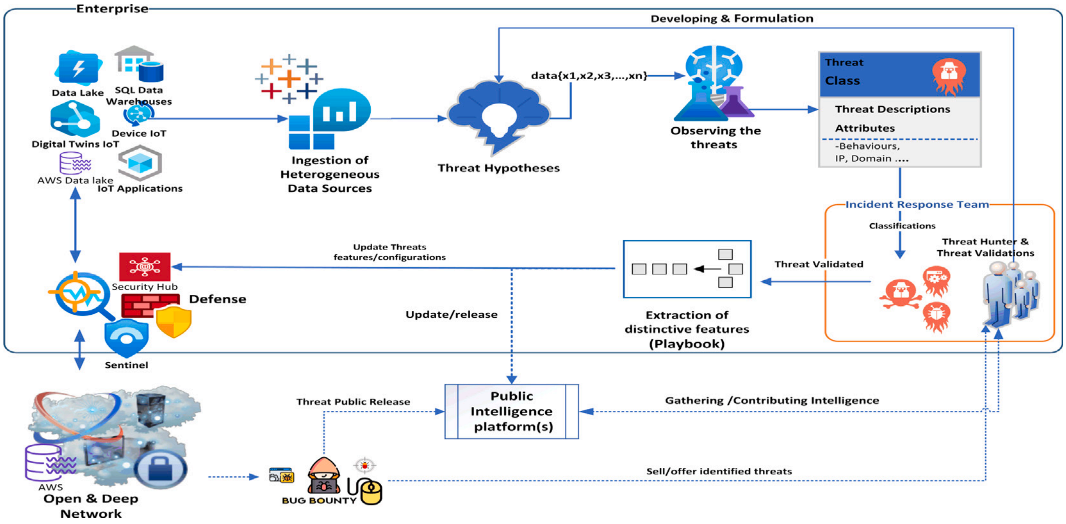

Fig. 2. Systematic process of adaptive threat hunting.

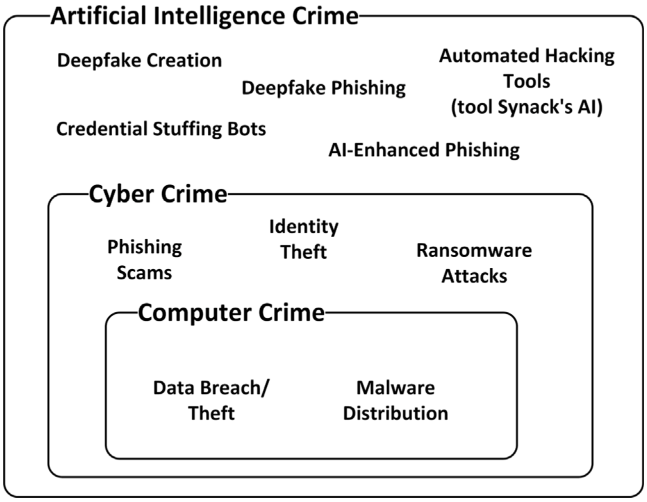

Fig. 3. The shift in computer crime towards ICT and AI technologies.

to learn, including ML, DL, reinforcement learning (RL), SVMs, and genetic algorithms, brings with it unforeseen repercussions, notably making criminal activities more streamlined and effective (Kaloudi and Li, 2020).

Given the increasing complexity and intelligence of attacks, now further driven by AI, the initial stages of threat hypothesis formulation need to be meticulously crafted. The generic threat hunting methodologies can be categorized as follows:

Analytics-Driven: This method involves coupling machine learning with User and Entity Behavior Analytics (UEBA) to generate aggregated risk scores, which can then be used as the basis for ranking hunting hypotheses.

Situational-Awareness Driven: This approach includes analyzing critical assets (Crown Jewel analysis), conducting enterprise risk assessments, and identifying company- or employee-level trends.

Intelligence-Driven: This strategy utilizes threat intelligence reports and feeds, malware analysis, and vulnerability scans to inform the hunting process.

2.4. IBM threat hunting methodology

IBM (2023) recognizes three different threat hunting methodologies, namely: *structured*; *unstructured*; and *situational*. Structured hunting is grounded in the detection of Indicators of Attack (IoA) and TTPs, employing CTI frameworks like the Lockheed Martin Cyber Kill Chain or the MITRE ATT&CK to anticipate and counteract threats (Straub, 2020). Conversely, unstructured hunting initiates from IoC, leading to broad investigations to unearth pre- and post-detection activity patterns without a predetermined framework. Finally, situational hunting, or entity-driven hunting, integrates both structured and unstructured elements based on internal analyses or trend observations, leveraging crowd-sourced data and telemetry to disclose novel TTPs and latent threats within organizational systems.

2.5. CrowdStrike threat hunting methodology

Drawing upon the methodologies espoused by cybersecurity leader CrowdStrike (2023), proactive threat hunting is predicated based on the hypothesis that adversaries may have already infiltrated the system, necessitating a vigilant search for signs of unusual behavior that could betray the presence of malicious activities. This vigilance is manifested in three core investigative strategies:

(i) hypothesis-driven investigations, which leverage the wealth of knowledge gained from crowdsourced attack data to unearth new adversary TTPs, thereby guiding the search for these malicious patterns within the organization's own digital environment;

(ii) investigations anchored in known IoCs or IoAs, employing tactical threat intelligence to spotlight IoCs and IoAs tied to emerging threats as the springboard for unearthing covert attacks or ongoing malevolent actions;

(iii) advanced analytics investigations that harness the formidable power of data analysis and ML to parse through extensive data troves in pursuit of anomalies that could indicate underlying threats with these detected discrepancies serving as the impetus for further scrutiny by adept analysts.

Emblematic of a harmonious blend of human acumen and cutting-edge security technology, these strategies exemplify a proactive, intelligence-led approach to securing organizational systems and data, as championed by CrowdStrike.

2.6. Resources for intelligence ontologies, knowledge-bases, and automation tools

In the realm of cybersecurity, the available tools, knowledge bases, and platforms for automation have significantly evolved. We have summarized some of the CTI concepts, platforms, and tools in Table 1. In this section, we discuss various important milestones in the development of publicly available tools and platforms.

Table 1
Overview of cyber threat intelligence: Concepts, Platforms, and Tools.

		Description		
Concepts	Cyber Threat Intelligence (CTI)	Information gathered and analyzed to understand and mitigate cyber threats.Tactics, Techniques, and Procedures (TTP)		
	Tactics, Techniques, and Procedures (TTP)	The behavior of threat actors, including their tactics (the why), techniques (the how), and procedures (the specific way)		
		Launched	Description	Key Features
Platforms and Frameworks	AlienVault Unified Security Management (USM)	**2007**	• AlienVault is a comprehensive security platform that combines multiple essential security capabilities within one console. • Now a part of AT&T Cybersecurity, it aims to simplify threat detection, incident response, and compliance management for IT teams of all sizes.	• The platform provides asset discovery, vulnerability assessment, intrusion detection, behavioral monitoring, and SIEM • A threat exchange network that enables collaborative defense with community-sourced threat data • Comes with built-in compliance reporting templates that help simplify meeting regulatory requirements.
	Malware Information Sharing Platform (MISP)	**2011**	• A platform for sharing threat intelligence among communities.	• Community-driven development • Swift exchange of information • Integration with the STIX format
	MITRE ATT&CK	**2016**	• A knowledge base of adversary tactics and techniques.	• Catalogs tactics, and techniques, linked to specific tools and actors • Employs public sources • Offers tools like Caldera, Caret, Car, ATT&CK Navigator, and TramIntroduction of sub-techniques
	Semi-Automated Cyber Threat Intelligence (ACT)	**2017**	• Enhances capabilities for CTI consumption, analysis, enrichment, and sharing.	• Aims to balance the efficiency of automated processes with the nuanced understanding that human analysts bring • Community-Driven Initiative • Integration into organizations' existing security operations
	OpenCTI	**2019**	• Platform for analyzing, sharing, and storing CTI	• Graph-based architecture • Supports linked data • Shares principles with the ACT project
		Description		
Relationships	MISP and STIX Format	• MISP as an alternative and integrator to STIX for threat intelligence sharing.		
	MITRE ATT&CK and Tactical CTI	• MITRE ATT&CK contributes to the practical application of CTI, focusing on tactics and techniques.		
	ACT Project & OpenCTI	• Both projects aim at enhancing CTI capabilities but differ in architectural and licensing approaches.		
		Features		
Tools Associated with MITRE ATT&CK	Caldera	• Automated adversary emulation system; uses ATT&CK model to test defenses against known tactics, techniques, and procedures. Enables identification of weaknesses in security postures.		
	Caret	• Provides capabilities for creating and managing analytic tests based on ATT&CK techniques. It helps in the validation of detection capabilities and improvement of threat hunting processes.		
	Car	• Cyber Analytic Repository; a knowledge base of analytics developed to detect adversary behaviors described in ATT&CK. Focuses on translating ATT&CK techniques into implementable detection strategies.		
	ATT&CK Navigator	• A web-based tool for visualizing and exploring the ATT&CK matrix; allows users to annotate and customize their views on ATT&CK tactics and techniques. Facilitates the planning of defenses and the identification of coverage gaps.		
	Tram	• Threat Report ATT&CK Mapping; automates the mapping of textual threat reports to ATT&CK techniques. Aids in quickly associating threat reporting with the relevant ATT&CK framework components.		

Initiated in 2011, the Malware Information Sharing Platform (MISP) project transformed into the MISP Threat Sharing Platform (Wagner et al., 2016). This transformation underscores its progression into a community-driven initiative focused on the development of the platform and the sharing of threat intelligence. MISP, accessible on GitHub, is renowned for its emphasis on the swift exchange of information. Its structure, closely integrated with its platform, positions MISP as a viable alternative to the STIX format, although it also supports exporting data to STIX.

The Structured Threat Information eXpression (STIX™), developed by the OASIS CTI Technical Committee, serves as a pivotal language and serialization format for exchanging CTI among organizations. Its core is an ontology representing the domain's concepts and their interrelations, which facilitates a shared understanding of, and actions against, cyber threats by encompassing objects like indicators, incidents, and threat actors. Predominantly serialized in JSON for its compatibility and readability, STIX's design is both expansive and adaptable, ensuring that the exchange of CTI remains consistent, machine-readable, and evolves with the cyber threat landscape, thus enhancing prevention and mitigation efforts.

Another milestone in the cybersecurity landscape was the publication of the MITRE ATT&CK framework in 2016. This online knowledge base catalogs adversary tactics and techniques, linking them to specific tools and threat actors based on empirical observations. The content is curated by the ATT&CK team and relies exclusively on public sources. MITRE further enhances the utility of this knowledge base by offering a suite of tools, including Caldera, Caret, Car, ATT&CK Navigator, and Tram. These tools empower users to leverage the extensive information contained in the ATT&CK framework. The introduction of sub-techniques and detailed procedures marks a significant advancement in connecting practical indicators with broader strategic models, paving the way for future research in tactical CTI (Wang et al., 2023).

The Semi-Automated Cyber Threat Intelligence (ACT) project (Bromander et al., 2021), launched in 2017, represents another step forward, offering capabilities for the consumption, analysis, enrichment, and sharing of CTI. This initiative has been instrumental in the development of the ACT platform. Following in 2019, OpenCTI was introduced, embodying principles similar to those of the ACT project. OpenCTI distinguishes itself with a graph-based architecture that facilitates the robust querying and integration of diverse data sources, underpinned by a data model that supports linked data. Notably, OpenCTI's choice of an open-source license marks a departure from the ACT platform's approach.

2.7. Threat hunting datasets

In this section, we introduce several well-known datasets frequently utilized by researchers in the field of threat hunting.

Security datasets (AWS, Linux, Windows), led by the open threat research forge (OTRF) community, are a vital platform for cybersecurity, offering a vast collection of curated datasets aligned with the MITRE ATT&CK framework, empowering professionals to refine detection and response strategies. With an open-source approach, it fosters collaboration, accelerates innovation, and democratizes access to resources. Detailed documentation and use cases support users in navigating dataset complexities, making it invaluable for training cybersecurity professionals and students.

Mossé Cybersecurity Institute offers a GitHub repository of multiple datasets specifically designed for practicing threat hunting for educational purposes. Answers to practical problems associated with the first dataset are available to assist learners. The datasets aim to enable users to determine which devices have been compromised, thereby offering a practical approach to understanding and mitigating cyber threats.

Awesome-threat-detection is a curated GitHub repository, which includes a wide array of tools, configuration guides, network monitoring resources, fingerprinting tools, and datasets for threat detection and hunting. It features links to various datasets, including the Mordor datasets which contain pre-recorded security events generated by simulated adversarial techniques, and other resources like SecRepo.com and the Boss of the SOC (BOTS) dataset versions 1–3, among others.

Awesome Threat Detection and Hunting library is a GitHub repository maintained by the threat hunting community, and compiles a significant list of resources related to threat detection, hunting, and intelligence. It provides links to various threat hunting rule sets for SIEM platforms like Splunk and ELK Stack, training documents, tools, datasets, frameworks, and other resources. This collection is particularly useful for those looking to dive deeper into the tools and methodologies employed in threat hunting.

Real-CyberSecurity-Datasets is a GitHub repository compiling various cybersecurity datasets that can be utilized for different security problems using ML and other methodologies. The datasets cover a wide range of topics including botnet and ransomware detection, malicious URLs, cloud security, and more, making it a rich resource for cybersecurity enthusiasts and professionals looking to delve into data analysis.

Awesome-Cybersecurity-Datasets is a curated GitHub repository including a variety of cybersecurity datasets related to phishing, passwords, malware, network traffic, and more. It includes links to resources like the Unified Host and Network Dataset, Comprehensive Multi-Source Cyber-Security Events, and datasets from the Canadian Institute for Cybersecurity, among others, making it an invaluable resource for those looking to study or develop cybersecurity solutions across a spectrum of challenges.

ARCS Datasets from Los Alamos National Laboratory is managed by Triad National Security, LLC, for the U.S. Department of Energy's NNSA. The ARCS datasets include comprehensive cybersecurity events, unified host and network datasets, and user-computer authentication associations over time. These datasets offer deep insights into the behavior of users and networks within a high-security environment and can serve as an excellent basis for developing and testing threat detection models.

IoT datasets are valuable resources for cybersecurity research, developed and maintained by various academic and research institutions. For instance, the **IoT-23 Dataset** and **Aposemat IoT-IDS Dataset** are provided by the Stratosphere Laboratory at the Czech Technical University (CTU), capturing network traffic from IoT devices to identify attack patterns. The **UNSW-NB15 Dataset** and **IoT Network Intrusion Dataset (TON_IoT)** are provided by the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) at the University of New South Wales (UNSW), containing comprehensive network traffic data including IoT devices for intrusion detection studies. Additionally, the **N-BaIoT Dataset** was created by researchers at Ben-Gurion University of the Negev, Israel, offering data on normal and botnet traffic from various IoT devices. These datasets enable a comprehensive analysis to understand IoT device behavior, identify vulnerabilities, and enhance security measures for IoT deployments.

CICDarknet2020 (Habibi Lashkari et al., 2020) is curated by the Canadian Institute for Cybersecurity (CIC). This dataset captures extensive network traffic from dark web sources, providing researchers with valuable insights into illicit online activities and cyber threats. CICDarknet2020 includes data such as forum posts, transaction records, and other interactions within dark web environments. This enables the study of emerging threats, criminal behaviors, and trends in cybercrime.

CIC-Bell-DNS-EXF-2021 dataset and CIC-Bell-DNS 2021 dataset are two well-known DNS datasets, consist of DNS queries, responses, and metadata collected from various sources. Collected and shared by the collaboration between the Canadian Institute for Cybersecurity (CIC) and Bell Canada, these DNS datasets can be used to analyze domain resolution patterns and identify malicious domains as well as activities such as DNS tunneling and domain generation algorithms (DGAs). They can also be used to identify anomalies, track the spread of malware, and enhance security measures to protect against DNS-based attacks.

IDS datasets is the group of Intrusion Detection Systems (IDS) datasets that are collected and/or provided by the Canadian Institute. They consist of various types of network attacks. For example, Realistic IDS- DoS and spoofing attack in IoV (CICIoV2024) is a recent dataset focused on Denial of Service (DoS) and spoofing attacks specifically targeting the Internet of Vehicles (IoV). CICEV2023 DDoS attack is a dataset on DDoS attacks. Intrusion detection evaluation dataset is a comprehensive dataset covering various intrusion attacks, including DDoS, perfect to be used for benchmarking IDS performance.

3. Hypothesis formulation

The initial step in a threat hunting program involves identifying its overarching goal: What is the rationale for the hunt? This fundamental question must be addressed by aligning the strategic plan with the organization's business objectives. The next critical step is to establish a hypothesis to guide the entire process. The central query for hypothesis formation is: What specific threats are we targeting? Domain knowledge is crucial for generating relevant hypotheses.

Therefore, this section comprehensively discusses the hypothesis formulation to address research question **RQ2**. In threat hunting, developing a hypothesis is an integral part of the process that enables cybersecurity professionals to proactively identify and mitigate potential threats before they can cause harm. A hypothesis in this context is essentially an educated guess or theory that seeks to explain the presence of malicious activity within an organization's network based on available data, trends, and security intelligence. Here we present a formal mathematical method for systematically approaching ransomware cyber attacks.

Consider, as an example, a ransomware attack which involves initial lateral movement within the network, then the identification and exploitation of critical resources, culminating in data encryption and a ransom request. Our goal is to formulate the substantial actions to be considered in a threat hunting hypothesis to define a series of mathematical expressions to represent the attacker's behavior within a network. The requisite steps in this modeling process are:

Attacker Model: We presuppose that the attacker has fixed attributes and capabilities. Let $A_{\mathtt{exp}}$ represent the attacker's general capability in exploiting a vulnerability. Likewise, let $A_{\mathtt{inj}}$, $A_{\mathtt{exf}}$ and $A_{\mathtt{enc}}$ denote the attacker's capability of process injection, data exfiltration, and data encryption, respectively. The likelihood of a successful attack will depend upon these attributes.

Network Model: The computer network is represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes or vertices (computers, servers, etc.), and $\mathcal{E}$ is the set of edges representing communication paths.

Lateral Movement: Let $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ be the set of paths an attacker takes to perform lateral movement, where each $p_i = (v_{i1}, v_{i2}, \ldots, v_{im_i})$ is a sequence of nodes representing a particular path taken through the network. The probability of moving from node v_i to node v_{i+1} is represented as $P(v_{i+1} \mid v_i)$.

Learning Network Architectures: Define a knowledge matrix $K = [K_{ij}]$ where K_{ij} represents the attacker's knowledge about the connection between nodes v_i and v_j. Initially, $K_{ij} = 0$. As the attacker learns about the network, K_{ij} increases, representing improved understanding.

Stealth Maintenance: Let $S(v) \in [0, 1]$ be a stealth score for node v, where a higher value indicates greater stealth. The attacker's actions decrease $S(v)$, making detection more likely. Define a detection function $D(v) = 1 - S(v)$, where a higher $D(v)$ means higher detection risk.

Critical Resource Identification: Critical resources are nodes $v \in \mathcal{C} \subseteq \mathcal{V}$ that are identified based on their network value N_v and resource value R_v. The identification function $I(v)$ combines these values, such that $I(v) > \theta$ indicates a critical resource for some predefined threshold θ, i.e. $\mathcal{C} = \{c \in \mathcal{V} \mid I(c) > \theta\}$.

Exploiting Vulnerabilities: For each critical resource $c \in C$, the exploitation probability is denoted as $E(c)$, which is a function of the resource's vulnerability score $V(c)$ and the attacker's exploitation capability $A_{\mathtt{exp}}$.

Process Injection: Let $PI(c)$ represent the success of process injection into critical resource c, where $PI(c) = E(c) \cdot A_{\mathtt{inj}}$, and $A_{\mathtt{inj}}$ is the attacker's process injection capability.

Data Exfiltration: Define $EX(c)$ as the exfiltration function for critical resource c, where $EX(c) = PI(c) \cdot A_{\mathtt{exf}}$, with $A_{\mathtt{exf}}$ being the attacker's data exfiltration capability.

Data Encryption and Ransom Request: Finally, the encryption and ransom request can be represented as a function $R(c) = EX(c) \cdot A_{\mathtt{enc}}$, where $A_{\mathtt{enc}}$ represents the attacker's capability to encrypt data and make a ransom request.

3.1. Hidden states and observations hypothesis

Collectively, the hidden state variables representing all attack stages are given by:

$$(\mathcal{P}, K, \mathcal{V}, \mathcal{C}, E, PI, EX, R)\,. \tag{1}$$

This unstructured view may notionally be restructured into a set S of mutually exclusive hidden states that represent the possible distinct stages of an attack. Our hypothesis is in what stages of attack threat hunters can observe the hidden states of the possible attack; for such a hypothesis we formulate the following notations:

Hidden states (attacker actions): $S = \{s_1, s_2, \ldots, s_N\}$.

Observable effects/actions: $O = \{v_1, v_2, \ldots, v_M\}$.

Observable effects are specific to each state, such as unusual network traffic (for lateral movement), access pattern anomalies (for learning network architectures), etc.

State transition probabilities: $A = \{a_{ij}\}$, where $a_{ij} = P(q_{t+1} = s_j \mid q_t = s_i)$.

Transition probabilities capture the likelihood of moving from one attack stage to another. These probabilities are crucial for understanding the attacker's progression through the network.

Observation probabilities: $B = \{b_j(k)\}$, where $b_j(k) = P(o_t = v_k \mid q_t = s_j)$ for continuous observations, often modeled with a probability density function.

Given the continuous nature of network data, we use a probability density function (e.g. Gaussian) to model the observation probabilities with parameters specific to the observations related to each attack stage.

Initial state distribution: $\boldsymbol{\pi} = \{\pi_i\}$, where $\pi_i = P(q_1 = s_i)$.

The initial state distribution reflects the likelihood of starting in a particular attack stage. Typically, the initial stage would be 'Lateral Movement'.

The continuous HMM is represented by the tuple $\lambda = (A, B, \boldsymbol{\pi})$. The objective is to compute the probability of a sequence of observations $\langle o_1, o_2, \ldots, o_T \rangle$ given the model λ, which can be computed using the forward algorithm for continuous observations.

Hypothesis on prediction and update. The prediction of the attacker's next step involves updating the state probabilities based on observed actions using the forward–backward algorithm or Viterbi algorithm for the most likely path of hidden states.

Forward algorithm (Continuous observations). Given a sequence of observations $\langle o_1, o_2, \ldots, o_T \rangle$, the forward probability $\alpha_t(i)$ is the joint probability of the partial observation sequence up to time t with state s_i at time t, given the model λ:

$$\alpha_t(i) = P(\langle o_1, o_2, \ldots, o_t], q_t = s_i \mid \lambda) \tag{2}$$

For continuous observations, $b_j(k)$ is typically defined by a probability density function, such as a Gaussian, which depends on the parameters (e.g., mean, variance) associated with observation v_k in state s_j.

The development of hypotheses is crucial in defining the scope of threat hunting activities. In our analysis, we sought to map out the potential behaviors of ransomware actors, from the initial stages of discovery and lateral movements to the act of demanding ransom from victims. For effective threat hunting, it is essential to establish clear objectives, whether it involves identifying unknown threats or leveraging known IoCs to track down specific threats. For instance, the observation of processes transmitting frequent beacons (Abu Talib et al., 2022) — whether through stochastic or deterministic behaviors within an organization — serves as a key point of analysis. It is important to recognize that not all beaconing activities are indicative of malicious intent. However, this behavior can assist threat hunters or incident response teams in classifying data within their playbooks as either malicious or benign.

By systematically formulating and examining these behaviors and employing a data-driven approach, we can enhance the precision of threat detection and response.

3.2. Anomaly detection and threat hunting

Anomaly detection in the context of cybersecurity is crucial for identifying patterns of behavior or network traffic that significantly deviate from the norm, signaling potential security breaches or malware infections.

Feature Space: $X = \{x_1, x_2, \ldots, x_n\}$ represents the feature space, where each x_i represents a vector of features extracted from system or network behavior, relevant to identifying potential cyber threats.

Normal Behavior Model: Define a model M that represents the normal behavior within the system, constructed using historical data of the system's operation. This model can be a statistical or machine learning model, such as a multivariate Gaussian model with mean μ and covariance matrix Σ estimated from the data.

Anomaly Score Function: An anomaly score function $a(x_i)$ quantifies the deviation of a data point x_i from the normal behavior model M. For a Gaussian model, the Mahalanobis distance can be used:

$$a(x_i) = \sqrt{(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)} \tag{3}$$

Threshold and Detection Rule: A threshold θ is defined to classify a data point as normal or anomalous. The detection rule is:

$$\text{Anomaly}(x_i) = \begin{cases} 1 & \text{if } a(x_i) > \theta \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

3.3. Threat hunting indicator function

The indicator function $t(x_i)$ determines whether an anomaly x_i warrants further investigation as a potential threat:

$$t(x_i) = \begin{cases} 1 & \text{if } c(x_i) \times a(x_i) > \tau \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where τ is a threshold indicating the level of concern required to trigger a threat hunting investigation. This threshold is determined based on the security posture and risk tolerance of the organization.

Defining a threshold for anomaly detection: In the context of detecting ransomware cyberattacks via anomaly detection, the determination of the threshold τ is crucial for balancing sensitivity and specificity, aiming to minimize false negatives (missing actual attacks) while controlling false positives (misidentifying normal behavior as attacks). The threshold can be mathematically formalized as follows:

Given a collection of behavioral metrics $X = (x_1, x_2, \dots, x_n)$ representing the normal operational state of a system and anomaly scores $\Gamma = (a(x_1), a(x_2), \dots, a(x_n))$ calculated from these metrics, the threshold τ is defined to maximize the detection of ransomware activities without overwhelming the system with false alarms.

The optimal threshold τ^* (offline learning data) can be defined as:

$$\tau^* = \arg\min_{\tau} \{\mu \cdot \text{FPR}(\tau) + (1 - \mu) \cdot [1 - \text{TPR}(\tau)]\} \tag{6}$$

where

FPR(τ) is the false positive rate at threshold τ, TPR(τ) is the true positive rate (sensitivity) at threshold τ, and μ is a weighting factor that balances the cost of false positives against missing true positives (ransomware attacks), reflecting the criticality of the assets under protection and the operational impact of responding to false alarms.

3.4. Iterative threat hunting and anomaly detection

The threat hunting process is iterative, with the results of investigations I feeding back into refining both the anomaly detection model M and the contextual analysis function $c(x_i)$:

$$(M', c') = f(M, c, I) \tag{7}$$

where M' and c' are the updated model and contextual analysis function, respectively, and f represents the update mechanism based on investigative findings I.

3.5. Playbooks and threat detection

Hypotheses in threat hunting can lead to the identification of specific threat behaviors, attack vectors, or security vulnerabilities that an organization might face. This process involves analyzing potential or emerging threats and formulating assumptions about how these threats could manifest within an organization's network. The development and implementation process of a collection of standardized procedures and guidelines is known as a "*playbook*" (Schlette et al., 2023).

Playbooks have emerged as a cornerstone for enhancing organizational defense mechanisms against cyber threats. A cybersecurity playbook is essentially a comprehensive manual that delineates a step-by-step approach for IT and security professionals to detect, respond to, and remediate various cyber threats and vulnerabilities. This manual serves not only as a procedural guide but also as a strategic document that outlines the identification of signs of malicious activities, analysis of potential threats, and the implementation of effective countermeasures to mitigate risks. The primary goal of these playbooks is to ensure that response teams can act swiftly and efficiently based on predefined protocols, tailored to address specific types of cyberattacks or security incidents. This systematic approach facilitates the standardization of threat response strategies, reduces response times, and, consequently, improves the overall security posture of organizations (Schlette et al., 2021a, 2023; Rizvi et al., 2022).

One of the most notable contributions to the cybersecurity domain, specifically in the realm of threat hunting, is the ThreatHunter-Playbook, hosted on GitHub by the OTRF community (Rodriguez and Rodriguez, 2022). It is an extensive collection of actionable intelligence and methodologies aimed at uncovering malicious activities within network environments. The playbook meticulously outlines various techniques employed by adversaries, closely aligning with the principles of the MITRE ATT&CK framework.

4. Review of threat hunting approaches

In this section, we address research question **RQ3** by conducting a systematic review and critical evaluation of the existing literature concerning threat hunting methodologies, with a particular emphasis on approaches integrating ML techniques. The use of ML significantly enhances cybersecurity, facilitating a more analytical and efficient defense mechanism with less reliance on time and human effort. Data play a crucial role in making informed decisions and devising strategic approaches that bolster the success of ML methods. Efforts to incorporate ML into threat hunting categorize solutions into five primary types: (i) *Supervised Machine Learning*; (ii) *Unsupervised Machine Learning* (iii) *Reasoning techniques*; (iv) *Graph-based approaches*; (v) *Rule-based approaches*. Additionally, we examine other approaches such as statistical methods and behavioral analytics, and their underlying principles, techniques, and methodological frameworks (see Fig. 4). By evaluating these diverse strategies, our objective is to elucidate their respective aims, the intricacies of their technical execution, and their effectiveness in identifying and mitigating cyber threats. Through this comprehensive analysis, we seek to identify gaps in the current state of research and suggest directions for future investigations, thereby contributing to the advancement of robust, intelligent Threat Hunting paradigms.

4.1. Supervised machine learning

Among the papers selected for our SLR, 19 of the selected papers apply supervised machine learning to detect the threats shown in Table 2. This section reviews supervised machine learning-based techniques for threat detection. In this section, we examine threat detection techniques based on supervised machine learning. We examine the method approach, the datasets used, as well as their strengths and weaknesses.

OpCodes (HaddadPajouh et al., 2018): This study presents a deep-learning-based approach for detecting malware in IoT devices, utilizing a deep recurrent neural etwork (RNN) framework with a focus on long short-term memory (LSTM) models. The methodology involves extracting operation codes (OpCodes) from ARM-based IoT applications to form a dataset comprising both malware and benign software samples. Through an innovative feature selection process and the implementation of LSTM networks, the approach effectively analyzes the sequential data of OpCode patterns to identify malware activities.

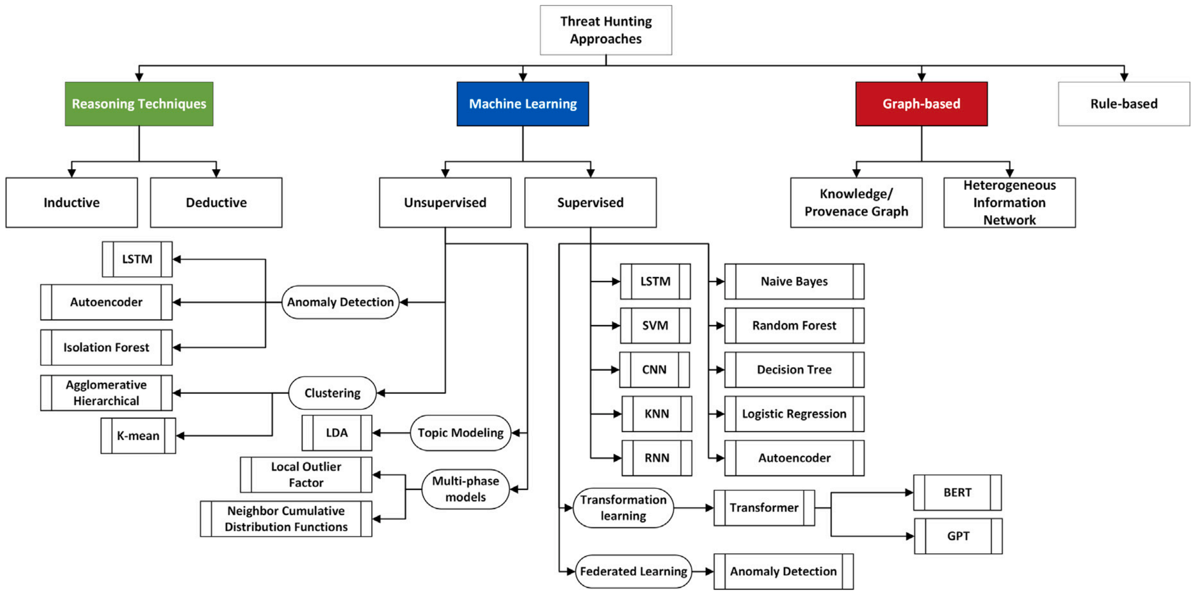

Fig. 4. Classification of threat hunting approaches.

The proposed system was trained and evaluated on a dataset of 281 malware and 270 benign samples, with further validation using 100 unseen malware samples. The findings reveal that a specific configuration of the LSTM model, characterized by a 2-layer architecture, achieves an impressive 98.18% accuracy in detecting new malware samples, significantly outperforming traditional machine learning classifiers. This research marks a pivotal step towards enhancing IoT security through the application of deep learning techniques, demonstrating the LSTM model's superior capability in learning and identifying complex malware behaviors from OpCode sequences.

The approach utilizes supervised learning, as it involves training the LSTM model on a labeled dataset that includes both malware (negative) and benign (positive) samples. Through the training process, the model learns to distinguish between the patterns associated with malware and those associated with benign software, allowing it to accurately classify new, unseen samples.

DRTHIS (Homayoun et al., 2019): This paper introduces a deep ransomware threat hunting and intelligence system (DRTHIS), specifically designed for deployment at the fog computing layer. The term "fog computing" refers to a computing strategy designed to bring the advantages of cloud computing closer to end-users and their IoT devices (Costa et al., 2022). The DRTHIS system employs advanced deep learning techniques, specifically LSTM and CNN, to offer a robust solution for identifying and classifying ransomware threats. By analyzing sequences of system actions and calls during the initial execution stages of applications, DRTHIS effectively differentiates between malicious ransomware and benign software. Furthermore, it precisely identifies various ransomware families, ensuring high precision in threat detection.

Tested on a comprehensive dataset, including 220 samples each from major ransomware families such as Locky, Cerber, and TeslaCrypt, alongside benign applications, DRTHIS achieved an impressive F-score of 99.6% for ransomware detection and demonstrated capabilities in recognizing previously unseen ransomware samples from new families. This approach applies in the fog computing domain, employing supervised deep learning methods to process and learn from complex data patterns inherent to ransomware behavior. Through the application of LSTM networks, the system adeptly handles the sequential nature of system call data, while CNNs contribute to extracting relevant features crucial for classification tasks.

Long-term and short-term dependency data (Yazdinejad et al., 2023): This study introduces a parallel ensemble machine learning model designed for accurate threat hunting in IIoT edge devices. By integrating several state-of-the-art classifiers, including decision trees (DTs), support vector machines (SVMs), logistic regression (LR), and random forest (RF), the model adeptly classifies multi-class anomalies through the synergistic use of multi-class AdaBoost and majority voting mechanisms. This ensemble approach, characterized by its parallel processing capabilities, significantly enhances the model's efficiency and accuracy in detecting anomalies across a diverse array of IIoT edge devices.

The model was rigorously evaluated using two real-world IIoT datasets, the Gas Pipeline (GP) and Secure Water Treatment (SWaT), both characterized by their imbalanced nature and containing both long-term and short-term dependency data. The findings demonstrate remarkable performance, with the proposed model achieving detection accuracies of 99.3% and 99.7% on the GP and SWaT datasets, respectively.

The core of the approach is not limited to conventional machine learning techniques; instead, it exploits the combined strengths of multiple classifiers within an ensemble framework to address the challenges of accurate anomaly detection in IIoT environments. This innovative method represents a significant leap forward in securing IIoT edge devices, offering a scalable and effective solution for the early detection and classification of potential cyber threats. Through this research, the authors contribute a valuable tool for enhancing the security posture of IIoT systems, underscoring the critical importance of advanced machine learning strategies in defending against the increasingly sophisticated landscape of cyber threats.

While the focus of this paper focus is predominantly on leveraging deep learning techniques for anomaly detection, its innovative approach in addressing dataset imbalances and utilizing the LSTM and AE architectures for learning and feature reduction underlines a significant advancement in applying machine learning methodologies to cybersecurity within the IIoT domain. The success of this model sets a new benchmark for future research in the field of IIoT cybersecurity, particularly in developing robust, efficient models for threat detection and prevention.

ICS MITRE ATT&CK (Arafune et al., 2022a): This paper proposes a novel framework aimed at enhancing the cybersecurity posture of Industrial Control Systems (ICS) through automated threat hunting. The methodology leverages the ICS MITRE ATT&CK framework to systematically identify and analyze TTPs utilized by adversaries, thereby

Table 2
Supervised machine learning application in threat hunting.

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
HaddadPajouh et al. (2018)	Deep RNN with LSTM	*ARM-based IoT applications dataset*: includes 281 malware, 270 benign, 100 unseen malware samples	Utilizes OpCode patterns from ARM-based IoT applications for malware activity identification	Dataset is small, method evaluation using solely accuracy is insufficient and unreliable.
Homayoun et al. (2019)	LSTM and CNN	*Ransomware dataset*: comprises 220 samples, including benign applications and ransomware from families such as Locky, Cerber, TeslaCrypt, CryptoWall, TorrentLocker, and Sage	Focuses on the fog computing layer, employs advanced deep learning for ransomware detection and classification	A limited dataset was used in the study, which may not represent real-world scenarios. It focused on F-Measure and lacked comparison with other models and techniques.
Yazdinejad et al. (2023)	DT, SVM, LR, RF, Auto-encoder architecture with Multi-class AdaBoost and majority voting	*The datasets Gas Pipeline (274,628 samples) and Secure Water Treatment (1,048,576 samples)*: contain data collected from the water treatment system, actuators, and sensors, collected in normal and attack situations	Addresses dataset imbalances with an ensemble deep learning model, combines LSTM and AE for anomaly detection in IIoT edge devices	Complexity and hyperparameter dependence may restrict its practical deployment and require considerable tuning.
Arafune et al. (2022a)	Supervised ML with SVM classifier, ICS MITRE ATT&CK framework	Sample datasets derived from MITRE ATT&CK for ICS	Leverages ICS MITRE ATT&CK for automated threat hunting, emphasizes automated detection and prediction of attacks	It relies on multiple technologies may introduce implementation challenges and require substantial maintenance.
Homayoun et al. (2020)	Sequential pattern mining, ML models (J48, RF, Bagging, MLP)	*Ransomware dataset*: 517 Locky, 535 Cerber, 572 TeslaCrypt ransomware samples, 220 benign applications	Employs sequential pattern mining for ransomware detection and classification, focusing on system activity logs	The limited dataset size and focus on specific ransomware families may hinder the generalizability and long-term effectiveness of the proposed method.
Aghamohammadpour et al. (2023)	Attribute-based method, DODAF integration	Not applicable	Uses DODAF for designing a threat hunting system in ICS, focuses on architectural design without specific ML models	Lacks real-world implementation or case studies to demonstrate its practical effectiveness. Qualitative evaluation based on expert judgments may introduce bias into the assessment process.
Yazdinejad et al. (2022)	Federated learning with various anomaly detection algorithms	Synthetic data, use smart factory applications	Utilizes federated learning for decentralized anomaly detection in IIoT, emphasizes privacy and reduced bandwidth usage	Clustering algorithms, network characteristics and data distribution differences across clusters in federated learning may introduce bias, impacting the final performance.
Abdel-Basset et al. (2022)	Multiscale convolutions, GRU-based autoencoder	*The ToN_IoT and LITNET2020 datasets*: ToN_IoT contains labeled IoT and IIoT data from 9 classes of IoT traffic, and LITNET2020 contains 85 network flow features corresponding to 12 attack types.	Combines multiscale convolutions and GRU for cyber threat detection in ICPS, deployed as a microservice	Fails to exploit widely available unlabeled data, differential privacy may negatively impact performance, and the system assumes all participants are trusted.
Farooq and Otaibi (2018)	Various ML algorithms (K-Means, DBSCAN, BIRCH, OCSVM, RF)	Synthetic data Microsoft Sysmon tool, focus on logs (Event ID: 1) collected locally on multiple enterprise hosts	Explores different ML techniques for reducing false positives in SOCs, focuses on specific cybersecurity scenarios	Lacks dataset information such as size, attack rate, and balance rate, essential for evaluating ML models. It also misses empirical validation with common metrics like accuracy, precision, recall, and F1-score for each model.
Bibi et al. (2023)	ConvLSTM2D model	Dataset with 21 million instances of attack patterns and threat vectors	Uses ConvLSTM2D for detecting multi-vector IIoT threats, emphasizes CUDA for efficient data processing	The computational complexity and the trade-off between detection accuracy and speed efficiency.
Shin et al. (2021)	NLP and supervised ML techniques	Real-world data from Twitter	Automates malware IOC extraction from tweets, employs NLP and ML for early threat detection	Potential data poisoning attacks; the proposed method is susceptible to false information being introduced intentionally.
Li et al. (2023a)	Transformer models, bi-directional LSTM	Open-source datasets (HDFS, OpenStack, PageRank, BGL logs)	Combines transformer models and bi-directional LSTM for detecting APT sequences and predicting attack paths	It incurs higher computational costs, particularly due to the use of transformer and bi-directional LSTM models. This can result in higher runtime overheads, making the method less efficient in real-time scenarios where quick detection and response are crucial.

(*continued on next page*)

Table 2 (*continued*).

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
Jahromi et al. (2020)	Stacked LSTM with pre-training regularization	Datasets with Windows Ransomware, IoT malware, and Android malware samples	Modifies stacked LSTM to avoid random initialization, focuses on rapid and accurate malware detection	Its reliance on careful tuning can involve significant experimentation, which may hinder the method's scalability and practicality.
Janjua et al. (2020)	AdaBoost, NB, LR, KNN, SVM	TWOS dataset	Analyzes email communications to detect insider threats, highlights the effectiveness of AdaBoost	The small size of the data presents a limitation, potentially affecting the generalizability and robustness of the results.
Pal et al. (2023)	LSTM and GRU with attention, ensemble techniques	CMU CERT insider threat datasets	Employs deep learning with attention mechanism for detecting insider threats, uses ensemble for classification	The method could miss detecting longer-term threats due to training the underlying model with single-day activity sequences.
Alsaheel et al. (2021)	Causality analysis, NLP, sequence-based ML	Dataset from executing ten real-world APT attacks	Automates construction of attack narratives from system audit logs, leverages causality analysis and ML	The method may miss attacks that mimic normal behavior patterns.
Kumar et al. (2023)	LSTM-VAE, Bi-GRU with sigmoid and softmax functions	ON_IoT datasets	Combines LSTM-VAE for feature extraction and Bi-GRU for detection and identification in maritime transportation systems, achieving up to 99% accuracy	The method uses LSTM-VAE and Bi-GRU-based schemes, requiring significant computational resources.
Shen and Stringhini (2019)	Cosine Similarity based as the distance metric to quantify the temporal embedding changes at time t in the latent space	Collected by authors., 190 million security events collected from tens of millions unique machines per day (102 consecutive weeks of data between December 1, 2016, and November 08, 2018)	Build an appropriate word embedding to effectively model and monitor the evolution of cyber attack events	The method relies a dataset of pre-labeled security events, which can result in missing unknown attacks such as zero-day vulnerability-based attacks.
Hemberg et al. (2024)	BERT	BRON dataset contains hundreds of thousands of nodes from D3FEND, CAR, Engage, ATT&CK, CAPEC, CWE	It demonstrates the utility of the enhanced BRON graph for hypothesis-driven cyber hunting and machine learning-enhanced simulations. The model leverages new data sources and machine learning to infer novel relationships, improving cyber defense strategies.	Limitations include data quality and bias, dynamic data nature, and constraints in the machine learning methodology, which require further refinement.
Liu et al. (2022)	Transformer embedding approach	1800 threat reports collected from Unit42, AlientVault	By adopting a Transformer-based architecture, ATHRNN enhances the semantic representations of threat reports, achieving state-of-the-art performance in extracting techniques. However, limitations include the incompleteness of the ATT&CK framework and the coarse granularity of extracted techniques.	Future work will focus on refining the ATT&CK set, extracting sub-techniques, and integrating regularization to improve hierarchical extraction.

automating the generation and validation of hypotheses regarding potential cyber threats. The framework comprises two principal components. The first involves the automatic detection of adversarial TTPs through a centralized threat hunting platform, which interfaces with the MITRE ATT&CK framework to discern potential threats based on observed network activities. The second component utilizes a supervised machine learning approach to predict future attack strategies, enabling proactive countermeasures.

Key results demonstrate the framework's capability to accurately identify network attacks, generate insightful hypotheses for threat hunters based on TTPs, and employ a machine learning classifier to foresee attackers' next moves. Through the creation of a detailed proof-of-concept, including sample datasets derived from the MITRE ATT&CK for ICS, the paper showcases the system's efficacy in real-world attack scenarios.

While the study heavily emphasizes the use of a supervised machine learning model, particularly an SVM classifier, for the prediction of future attack behaviors, it is equally rooted in the automation of the threat hunting process itself. The approach systematically organizes threat intelligence and employs algorithmic analysis to automatically detect and respond to cyber threats, significantly reducing reliance on human intervention and minimizing human error. This automated mechanism, independent of the machine learning aspect, constitutes a significant portion of the research, showcasing an innovative blend of cybersecurity frameworks and algorithmic processing to fortify ICS networks against sophisticated cyber threats.

Sequential pattern mining (Homayoun et al., 2020): This study presents an approach to ransomware detection and classification using sequential pattern mining techniques. Focusing on the analysis of system activity logs, the study identifies maximal frequent patterns (MFPs) within logs of known ransomware families, such as Locky, Cerber, and TeslaCrypt, alongside benign applications. These patterns serve as distinct features that are leveraged in machine learning models for efficient classification. The dataset consists of 517 Locky, 535 Cerber, and 572 TeslaCrypt ransomware samples, and 220 benign applications, facilitating a comprehensive analysis. By applying sequential pattern mining to extract meaningful patterns and using machine learning algorithms, including J48, Random Forest, Bagging, and MLP, the study achieves a remarkable 99% accuracy in distinguishing ransomware from benign software and 96.5% accuracy in classifying ransomware into specific families.

The core approach of this research is not solely reliant on conventional machine learning techniques but significantly incorporates sequential pattern mining to discover and utilize MFPs as features for classification.

D3FEND (Aghamohammadpour et al., 2023): This paper introduces a novel architecture for developing automated threat hunting systems within ICS using the Department of Defense Architecture Framework (DoDAF). This research leverages DoDAF to systematically design a threat hunting system that addresses the complex cybersecurity challenges inherent in ICS. By integrating MITRE's ATT&CK and D3FEND frameworks, the proposed architecture enhances its capability to detect, analyze, and mitigate cyber threats more effectively. The approach focuses on creating an attribute-based method to categorize and understand malicious threats by analyzing similarities between suspicious and known malicious events.

A significant part of the research involved evaluating the system's architecture using twelve essential quality attributes. This evaluation was conducted through a survey questionnaire approach, gathering insights from cybersecurity experts to assess the impact of these attributes on the development and effectiveness of the threat hunting process. Although the study does not directly employ machine learning methods within the architecture, it suggests an analytical approach to comparing threat categories and enhancing the understanding of cyber threats in ICS environments. The absence of a specific dataset for training or evaluating machine learning models highlights that the paper's primary contribution is towards the strategic architectural design and qualitative evaluation of a threat hunting system.

Block Hunter (Yazdinejad et al., 2022): This paper introduces the Block Hunter framework, designed to fortify cyber threat detection within the IIoT networks leveraging blockchain technology. The core framework lies in employing federated learning (FL) to establish a decentralized anomaly detection mechanism that enhances privacy preservation and reduces bandwidth usage without centralizing data. Block Hunter's architecture is distinguished by its cluster-based approach, facilitating efficient anomaly detection through collaborative learning among various machine learning models in a federated environment.

This setup not only aims to identify anomalous behavior with high accuracy but also addresses critical challenges such as training data scarcity and privacy concerns that are prevalent in IIoT applications. Although the research emphasizes federated learning as its primary approach, it acknowledges the integration of various anomaly detection algorithms, including clustering-based, statistical, subspace-based, classifier-based, and tree-based methods, to bolster the efficacy of the Block Hunter framework.

Fed-TH (Abdel-Basset et al., 2022): This study develops the federated threat-hunting (Fed-TH) deep learning framework optimized for detecting cyber threats in industrial cyber–physical systems (ICPS). Fed-TH integrates two key models: multiscale convolutions for capturing spatial data features; and a gated recurrent unit (GRU)-based autoencoder for temporal data analysis. This combination enables the framework to effectively process and analyze the complex data patterns inherent in ICPS, facilitating accurate threat detection.

Deployed as a microservice on edge servers within a container-based edge computing architecture, Fed-TH leverages federated learning to decentralize data processing, thereby enhancing privacy and reducing latency by eliminating the need to transmit sensitive data over the network. The effectiveness of Fed-TH is demonstrated through its application to two public benchmarks, where it achieved notable accuracy and F1 scores. By utilizing a federated learning approach, Fed-TH addresses key issues related to data privacy, latency, and computational resource optimization in the cybersecurity domain for ICPS. The deployment strategy, emphasizing microservice architecture on edge servers, underscores the practical applicability and scalability of this solution in industrial settings.

SOCs (Farooq and Otaibi, 2018): This paper delves into the critical evaluation and application of various ML algorithms to enhance threat detection capabilities within security operations centers (SOCs). Addressing the prevalent issue of high false-positive rates in cyber threat detection, the study systematically explores the use of different ML techniques, including clustering algorithms (K-Means, DBSCAN, BIRCH), One-Class SVM (OCSVM) for anomaly detection, and random forest for message classification, to analyze security logs and detect potential threats more accurately. The approach taken involves applying these ML algorithms to specific cybersecurity scenarios, such as analyzing internet traffic data for abnormal patterns, detecting anomalous process executions within Windows environments, and classifying enterprise mobile messaging data.

This methodological application aims to identify the most effective algorithms for reducing false positives and improving the overall efficiency of SOC operations. The study does not provide detailed empirical results nor does it list the specific datasets used but it implies the analysis of varied data relevant to each cyber threat detection scenario. The selection of algorithms is guided by their potential to model the data effectively and reduce false positives, showcasing the practical implications of ML in enhancing cyber threat detection capabilities.

2D Convolutional LSTM (Bibi et al., 2023): This study introduces an innovative approach for enhancing cybersecurity in distributed IIoT environments. Leveraging a 2D Convolutional LSTM (ConvLSTM2D) model, the study proposes a scalable and self-optimizing mechanism capable of addressing the dynamic and sophisticated nature of emerging threats in IIoT systems. This deep-learning-based solution utilizes the compute-unified device architecture (CUDA) to efficiently process spatial and temporal data, making it particularly effective against multi-vector IIoT threats.

The ConvLSTM2D model's performance was rigorously evaluated using a comprehensive dataset comprising 21 million instances of various attack patterns and threat vectors relevant to IIoT systems. The dataset included categories such as man-in-the-middle (MitM) attacks, denial of service (DoS), botnet malware, and reconnaissance activities. The results demonstrate the ConvLSTM2D model's superior detection accuracy compared to contemporary deep learning architectures and benchmark algorithms, although it presents a trade-off in terms of processing speed. Employing supervised learning techniques, the ConvLSTM2D model successfully classifies and identifies diverse cyber threats in IIoT environments.

#Twiti (Shin et al., 2021): This paper unveils a system designed to harness the power of social media, specifically Twitter, for CTI by automatically extracting malware IOCs from tweets. Employing a combination of natural language processing (NLP) and supervised machine learning techniques, Twiti identifies and classifies tweets potentially containing valuable IOC information, then meticulously extracts and organizes these IOCs from both the tweets and their linked content. The methodology encompasses a system that begins with collecting tweets based on cybersecurity-related keywords and user tracking, followed by a relevant tweet selection process using a sophisticated tweet classifier to filter out irrelevant information. Finally, the IOC extractor component employs pattern matching to identify and extract the IOCs, even handling defanged IOCs to ensure the integrity of the data collected.

Twiti's performance was evaluated using real-world data, demonstrating a high volume of accurate IOC extractions that surpassed existing threat intelligence systems in both precision and the uniqueness of the collected IOCs. Notably, Twiti managed to detect a significant portion of IOCs earlier than other public threat intelligence feeds, highlighting its potential for early threat detection and contribution to proactive cybersecurity measures.

DeepAG (Li et al., 2023a): This study introduces an innovative attack-graph framework designed for advanced threat detection and predictive analysis in cybersecurity. Leveraging the robust capabilities of transformer models and bi-directional LSTM (BiLSTM) networks, DeepAG analyzes system logs to detect APT sequences and predict potential attack paths with exceptional accuracy. This dual approach not only enhances the detection of malicious activities but also aids in the proactive identification of future threats by constructing visual attack graphs that represent possible attack strategies. DeepAG's methodology is distinguished by its use of transformer models to semantically analyze log sequences, transforming them into vectors for parallel processing. This reduces semantic information loss and the time costs associated with threat detection. Additionally, the framework employs a BiLSTM network that surpasses traditional approaches by efficiently locating anomaly points within system logs, thereby improving the accuracy of attack point confirmation.

Evaluated on open-source datasets comprising various system logs, DeepAG demonstrates a remarkable ability to accurately detect over 99% of more than 15,000 sequences, significantly outperforming baseline models. The bi-directional predictive model further enhances the framework's efficacy by improving baseline accuracy by 11.166% in threat location, showcasing DeepAG's potential in both detecting existing threats and predicting future attack paths. The experiments were conducted using open-source datasets of four different system logs: HDFS; OpenStack; PageRank; and BGL logs. These datasets comprise a variety of system logs that are used to evaluate the effectiveness of DeepAG in detecting APT sequences and predicting potential attack paths.

Capturing global and short-term dependencies (Jahromi et al., 2020): This paper introduces an innovative deep learning approach tailored to enhancing malware detection in cybersecurity applications. The study focuses on deploying an advanced, stacked LSTM network, uniquely modified to bypass the challenges associated with random initialization through a pre-training regularization method. This novel approach aims to capture both global and short-term dependencies in the data more effectively, crucial for identifying malware in safety and time-critical systems such as healthcare devices and military IoT applications. The methodology centers on utilizing the network for analyzing malware OpCode sequences, with a special emphasis on pre-training the model to ensure a more structured and informed learning process. This technique addresses the key challenge of achieving rapid and accurate malware detection without the computational overhead typically associated with deep recurrent neural networks.

The evaluation of the proposed method, conducted on datasets comprising Windows Ransomware, IoT malware, and Android malware samples, demonstrates a significant improvement in malware detection capabilities. The method achieved a remarkable detection accuracy of 99.1% for IoT malware samples, alongside an Area Under the Curve (AUC) of 0.985 and a Matthews Correlation Coefficient (MCC) of 0.95. These results underscore the efficacy of the proposed LSTM network modification over traditional methods, highlighting its potential to enhance the robustness and reliability of cyber threat detection systems.

TWOS (Janjua et al., 2020): This study investigates the efficacy of supervised ML algorithms in detecting potential insider threats within an organization, focusing on the analysis of email communications. Utilizing the TWOS dataset, which contains behavior traces of 24 users over five days, the research applies a range of machine learning algorithms — AdaBoost, naive Bayes (NB), logistic regression (LR), K-nearest neighbors (KNN), linear regression (LR), and SVM — to classify emails into normal and anomalous categories based on linguistic analysis. The approach involves preprocessing email logs through stemming, stop word removal, and tokenization, followed by the application of the aforementioned algorithms to identify potentially malicious emails.

Of the algorithms tested, AdaBoost emerged as the most effective, achieving a notable 98.3% accuracy rate and an AUC of 0.983 for the identification of malicious emails. This result highlights the superior performance of AdaBoost in classifying emails accurately and underscores the potential of supervised learning techniques in addressing the challenge of insider threats. The paper emphasizes the advantages of supervised learning in dealing with evolving threat concepts and identifying insider threats with limited labeled data.

CMU CERT insider threat (Pal et al., 2023): This paper introduces an ML-based approach to detect insider threats by analyzing user activity logs. Leveraging an ensemble of deep learning models, specifically stacked LSTM and GRU networks equipped with an attention mechanism, this method efficiently processes sequential daily user activities to identify potential insider threats. The approach is designed to highlight critical behavioral patterns and address the challenge of data imbalance in insider threat detection datasets. The methodology involves preprocessing the Carnegie Mellon University (CMU) computer emergency response team (CERT) insider threat dataset to generate single-day activity sequences, transforming these sequences into detailed action IDs, and implementing an equally weighted random sampling strategy to mitigate data imbalance.

The core of the detection framework employs LSTM and GRU models with attention layers to extract meaningful temporal features from activity logs, emphasizing segments indicative of malicious behavior. These features are then classified using ensemble techniques, combining AdaBoost and random forest classifiers for the final threat detection. The experimental evaluation on the CMU CERT insider threat datasets (versions 4.2, 5.2, and 6.2) demonstrates the approach's high efficacy, achieving an average AUC of 0.99 for versions 4.2 and 5.2, and 0.97 for version 6.2.

ATLAS (Alsaheel et al., 2021): This study introduces an innovative framework designed to automate the construction of attack narratives from system audit logs. By leveraging a novel combination of causality analysis, NLP, and sequence-based machine learning techniques, ATLAS effectively processes and analyzes system logs to identify and reconstruct the sequence of events that constitute a cyber attack. ATLAS operates through a three-phase approach: constructing a causal dependency graph from audit logs; generating timestamp-ordered event sequences representing attack and non-attack activities; and employing a sequence-based model to discern attack patterns within these sequences. This methodology enables the system to abstract complex attack behaviors into manageable sequences, facilitating the identification of critical attack steps and entities involved in the incident.

The evaluation of ATLAS involved executing ten real-world APT attacks in a controlled virtual environment, resulting in a dataset rich with both attack and non-attack activities. The system demonstrated high accuracy in recovering attack steps and constructing coherent attack narratives, achieving an average precision of 91.06%, recall of 97.29%, and an F1-score of 93.76%.

DLTIF (Kumar et al., 2023) This paper introduces the deep learning-driven cyber-threat intelligence modeling and identification framework (DLTIF) in IoT-enabled Maritime Transportation Systems (MTS). This framework aims to model CTI and identify specific threat types using deep learning techniques. DLTIF consists of three primary components: a deep feature extractor (DFE); a CTI-driven detection (CTIDD) scheme; and a CTI-attack type identification (CTIATI) scheme. The DFE scheme leverages an LSTM-based variational autoencoder (LSTM-VAE) to automatically extract latent threat patterns from network data. The CTIDD scheme, which utilizes a bi-directional GRU (Bi-GRU) combined with a sigmoid function, uses the output from the DFE scheme for threat detection. Finally, the CTIATI scheme is designed to identify the exact threat types using a Bi-GRU combined with a softmax function.

The framework was evaluated against five public datasets and demonstrated superior performance over seven state-of-the-art host intrusion detection systems. DLTIF achieved up to 99% accuracy, outperforming traditional and recent state-of-the-art approaches. The use of deep learning, specifically the combination of LSTM-VAE for feature extraction and Bi-GRU for detection and identification, enables DLTIF to effectively deal with the dynamic and sophisticated nature of cyber threats in IoT-enabled MTS.

ATTACK2VEC (Shen and Stringhini, 2019): This study presents an approach to analyzing the progression and transformation of cyberattacks over time. By employing temporal word embeddings, a technique inspired by NLP, the system with cosine similarity, dubbed ATTACK2VEC, transforms the landscape of cybersecurity event analysis. This method conceptualizes attack steps as 'words' and sequences of attacks as 'sentences', enabling a novel analysis of how these attack steps coalesce and evolve, akin to tracking changes in language usage over time.

The study leverages an extensive dataset comprising more than 190 million security alerts collected from a commercial intrusion prevention system (IPS) that spans two years. This datasets encapsulates a broad spectrum of security events, from routine port scans to specific exploitations of vulnerabilities identified by common vulnerabilities and exposures (CVEs), observed across tens of millions of unique machines.

Through ATTACK2VEC, the authors demonstrate the system's capability to monitor the emergence of new attack strategies and the associations between various attack steps, shedding light on the attackers' evolving tactics. In particular, the system successfully flagged the rise of a Mirai botnet variant targeting specific CVEs well before its formal identification within the cybersecurity community.

BRON (Hemberg et al., 2024): This novelty framework aims to use a transformer model which is a bidirectional relational knowledge from additional data sources. It demonstrates the utility of the enhanced BRON graph for hypothesis-driven cyber hunting and machine learning-enhanced simulations. The model leverages new data sources and machine learning to infer novel relationships, improving cyber defense strategies. Limitations include data quality and bias, dynamic data nature, and constraints in the machine learning methodology, which require further refinement.

ATHRNN (Liu et al., 2022): The study introduces the ATHRNN framework, which incorporates hierarchical dependencies and semantic information in ATT&CK extraction by encoding labels into matrices to learn relationships among reports and labels. By adopting a Transformer-based architecture, ATHRNN enhances the semantic representations of threat reports, achieving state-of-the-art performance in extracting techniques. However, limitations include the incompleteness of the ATT&CK framework and the coarse granularity of extracted techniques. Future work will focus on refining the ATT&CK set, extracting sub-techniques, and integrating regularization to improve hierarchical extraction.

LLM (Chang et al., 2023): The authors highlight the evolving role of large language models (LLMs) in semantic understanding, demonstrating their capabilities in interpreting language and its concepts, yet with limitations in perceiving semantic similarities among events. The survey found LLMs to be proficient in reasoning about causal and intentional relations and predicting future events with adequate context, but less effective in other relational types. Despite improvements, LLMs like GPT-4 still lag behind human performance in distinguishing meaningful phrases from nonsense, underscoring the need for further enhancement in semantic processing. Consequently, the improvement and optimization of LLMs in reasoning capabilities have significant implications for enhancing threat detection systems. By addressing limitations in abstract, multi-step, and domain-specific reasoning, LLMs can better identify and predict potential security threats in various contexts. Ongoing research to refine these models contributes to developing more sophisticated and accurate cybersecurity tools, improving the detection of complex cyber threats, and supporting more robust defense mechanisms against evolving digital vulnerabilities.

4.2. Unsupervised machine learning

This section reviews unsupervised machine learning-based techniques for threat detection as shown in Table 3. We divide them into three subgroups of techniques: (i) anomaly detection; (ii) clustering; (iii) topic modeling and unsupervised multi-phase models.

Anomaly detection approaches. Anomaly detection is a key technique in discovering unusual patterns, behaviors, or activities that diverge from established norms within a system or network. Since anomaly detection is intrinsically aligned with cyber threat detection, it plays a crucial role in detecting and enhancing the effectiveness of the threat detection landscape. Therefore, various methods and techniques have been meticulously developed and explored to maximize its potential.

LogAnomaly (Meng et al., 2019): The study is designed to represent a log stream as natural language sequences and then into log templates. It incorporates two phases: offline training and online detection. Utilizing LSTM networks, LogAnomaly learns from observed log sequences in the offline training phase. Logs that diverge significantly from predicted patterns, indicating unexpected behavior, are identified as anomalies during online detection.

LogUAD (Wang et al., 2022b): The model employs Word2Vec for log-based unsupervised anomaly detection. Before entering the anomaly detection algorithm, original log messages are transformed into Word2Vec vectors. The anomaly detection process uses K-means clustering along with predetermined thresholds.

Autoencoders & anomaly (Farzad and Gulliver, 2020): The model leverages isolation forest with two deep autoencoder networks for log message anomaly detection. While autoencoders manage training and feature extraction, isolation forest serves as the anomaly detection method for identifying unusual samples.

Clustering approaches. As a threat-hunting technique, clustering is used to find patterns, anomalies, or groups of similar entities in various types of datasets, such as network traffic, log files, or file characteristics.

LogCluster (Lin et al., 2016): The research creates log sequences and clusters them using agglomerative hierarchical clustering. Subsequently, representative log sequences from each cluster are extracted for further analysis. A knowledge base is consulted to determine if these log sequences have been observed previously and to identify potentially malicious logs. Any unseen log sequences are also manually examined to ensure they contain no threats.

LogKernel (Li et al., 2022b): The three-phase model follows the construction of a behavior provenance graph with graph kernel clustering. In the next step, the clustering results are assessed, and the most suspicious clusters are selected based on a threshold score. The threat score of each cluster is derived from a threat quantification formula that considers IP addresses, URLs, user activities, and sensitive information.

While some studies have employed straightforward clustering methodologies such as K-means or DBScan to cluster network event logs (Sharma and Parvat, 2013), to detect DDoS attacks (Al-mamory and Algelal, 2017), or to detect malware based on file attributes (Rosli et al., 2019), the evolving threat landscape requires more sophisticated methods. Advanced threat models necessitate adopting intricate techniques or multi-phase methodologies to ensure resilience against emerging and adaptive threats.

Topic modeling approaches. Topic modeling has also been used as a valuable tool in threat hunting for identifying patterns, themes, and anomalies within large volumes of textual data such as security logs, event logs, and incident reports.

CD-LDA (Satpathi et al., 2019): The study addresses the challenge of categorizing error messages from large distributed data center networks into distinct error events. This method leverages a unique approach to transform error events into episodes, treating them as documents in a document-term matrix. This transformation is achieved using non-parametric change-point detection techniques. Once the episodes are mapped to the document-term space, the latent dirichlet allocation (LDA) algorithm is employed for topic discovery. LDA helps in identifying underlying themes or topics within the error messages, facilitating the separation and classification of error events based on their shared characteristics.

Adams et al. (2018): The study introduces an automated approach to rank attack patterns within the common attack pattern enumeration

Table 3
Unsupervised machine learning application in threat hunting.

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
Meng et al. (2019)	Anomaly detection by NLP, LSTM	BGL dataset (4,747,963 logs) and HDFS dataset (11,175,629 logs) (Oliner and Stearley, 2007)	The model learns from observed log sequences in the offline training phase	There is a challenge in accurately approximating previously unseen logs due to the use of FT-Tree algorithm.
Wang et al. (2022b)	Anomaly detection by Word2Vec and K-means clustering	BGL dataset (4,747,963 logs) (Oliner and Stearley, 2007)	Original messages are vectorized before using K-mean clustering with predetermined thresholds	The method depends on the optimal dimensionality of word vectors generated by Word2Vec, which requires extensive tuning to ensure accurate anomaly detection.
Farzad and Gulliver (2020)	Anomaly detection by Autoencoder	BGL dataset (4,747,963 logs) (Oliner and Stearley, 2007), OpenStack (Du et al., 2017a) and Thunderbird (Oliner and Stearley, 2007)	Combine Autoencoder and Isolation Forest to identify unusual samples in huge datasets	The use of two deep Autoencoder networks increases the computational resources and time required for training and feature extraction.
Lin et al. (2016)	Agglomerative Hierarchical	Hadoop-based Big Data (Shang et al., 2013)	Log sequences from each cluster are extracted to determine potential malicious logs	The reliance on hierarchical clustering, which can be computationally expensive and less scalable especially for large-scale online service systems.
Li et al. (2022b)	Kernel clustering	DAPRA TC dataset	Scoring each clustered by considering IP addresses, URLs, user activities and sensitive information	The proposed method may not be able to detect attacks that do not use system call interfaces.
Satpathi et al. (2019)	Parametric change-point detection techniques	Virtual network function dataset (97 million raw syslog messages and 728 000 messages)	Categorizing error messages from large distributed data center networks into distinct error events	The method's accuracy is highly sensitive to the choice of hyperparameters.
Adams et al. (2018)	LDA	CAPEC dataset	Prioritizing and ranking attack patterns	Its heavy reliance on the quality and completeness of the textual descriptions in the CAPEC database of both the attack patterns and the system.
Gao et al. (2021a)	NLP and Unsupervised learning technique	DAPRA TC dataset	Automate theat hunting process by extracting structured threat behaviors from unstrutured OSCTI reports.	THREATRAPTOR's fuzzy search mode takes significantly longer execution times compared to its exact search mode, which can negatively impact the applicability of the approach.
Chen et al. (2022a)	Local Outlier Factor, Isolation Forest	APT3 dataset (8329 events) and Classic dataset (13 867 events)	Integration of a machine learning pipeline to detect cyber attack	Its inability to fully handle the challenge of imbalanced datasets.
Kayhan et al. (2023)	Autoencoders	Real data from VirusTotal behavior reports (50 000 entries)	Detect anomalous commands within Security Information and Event Management (SIEM) logs	It can only detect unique commands at the organizational level rather than at the user or business function level, potentially missing important context-specific anomalies.
Yousef et al. (2021)	Neighborhood Cumulative Distribution Function (NCDF) space	Simulated dataset and two real datasets (CICIDS2017 Sharafaldin et al., 2018 comprising 2,827,595 observations, 556,541 of which are malicious and CIRA-CIC-DoHBrw-2020 MontazeriShatoori et al., 2020 comprising comprises 1,139,362 observations, 249,836 of which are malicious)	Combine both manual review and algorithmic analysis to assign nuanced, continuous scores to data points without the need for training or labeling	The construction of the NCDF space is computationally intensive, which makes the method inefficient for detecting anomalies in large datasets.

and classification (CAPEC) datasets for a given system. It first extracts and analyzes the textual content of the attack dataset using latent dirichlet allocation (LDA) for topic modeling. Subsequently, the learned topic model from the CAPEC dataset is utilized to estimate a posterior distribution of topics relevant to the target system. By prioritizing and ranking attack patterns under the system's characteristics, the cybersecurity expert can detect threats more efficiently in the later stage.

Multi-phase model approaches. Unsupervised threat detection has seen the development of multi-phase model approaches that employ various techniques to progress through multiple phases, enhancing threat detection and knowledge integration.

THREATRAPTOR (Gao et al., 2021a): This study introduces a system designed to automate cyber threat hunting by leveraging OSCTI. THREATRAPTOR distinguishes itself with a specialized, unsupervised, lightweight NLP pipeline tailored for extracting structured threat behaviors from unstructured OSCTI reports, significantly outperforming traditional information extraction methods. The system employs a domain-specific query language, threat behavior query language (TBQL), which, coupled with an automated query synthesis mechanism, enables the concise and expressive querying of system audit logs for identifying malicious activities. Additionally, THREATRAPTOR features an efficient query execution engine that optimizes data storage and querying processes, markedly improving search efficiency across large datasets. The system's capability extends to a fuzzy search mode, enhancing the generality of threat hunting queries through inexact graph pattern matching.

Evaluation across a comprehensive set of attack cases demonstrates THREATRAPTOR's high accuracy in both threat behavior extraction and malicious activity detection, with notable efficiency in processing and querying operations. The system's use of an unsupervised NLP pipeline for threat behavior extraction from OSCTI text categorizes its learning approach as unsupervised. THREATRAPTOR's innovative approach to automated cyber threat hunting marks a significant advancement in cybersecurity practices, offering an efficient, accurate,

Table 4
Reasoning techniques application in threat hunting.

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
Narayanan et al. (2018)	Knowledge graph reasoner	Unified Cybersecurity Ontology (Syed et al., 2016)	Proposes a cognitive assistant based on open-source intelligence which has different terms for different audiences to early detect cybersecurity attacks.	However, the current method is limited by its reactive nature, which means that a response is generated only after threats have occurred and relies on the record of past attack patterns. Therefore, the study proposes using the dataset collected from the dark net and hacker communities to expand the record of attack patterns for better response when threats are detected.
Marin et al. (2020)	Causal reasoning (Kleinberg and Mishra, 2012) and logic programming (the Point Frequency function from Annotated Probabilistic Temporal logic Shakarian et al., 2011)	The authors collected 7824 posts with vulnerability mentions from 56 dark web forums and correlate them with cyber incidents to predict imminent cyber-attacks	Considers the socio-personal and technical indicators of enterprise attacks to predict imminent cyber incidents	The framework's deductive reasoning method is limited by its reliance on pre-established rules and patterns, potentially overlooking new or evolving attack strategies. Future work aims to integrate adaptive and learning-based approaches to improve predictive accuracy and adapt to emerging cyber threats.
Dritsoula et al. (2017)	Deductive reasoning with game theoretic approach	The authors simulated 3 fake news or malware apps scenarios	Takes into account the attacker's tradeoff to classify adverseries	The game-theoretic model's primary limitation is its assumption of rational behavior from attackers and defenders. Future work should consider scenarios where attackers may behave irrationally or unpredictably, enhancing the model's robustness and applicability.

and automated solution to the increasingly complex landscape of cyber threats.

Fuchikoma (Chen et al., 2022a): This study presents a system named Fuchikoma, which implements a sophisticated approach to cyber threat hunting through the deployment of various ML techniques, including anomaly detection algorithms such as local outlier factor, isolation forest, and DBScan, as well as graph algorithms aimed at identifying and investigating cyber threats effectively. Central to this methodology is the integration of a machine learning pipeline that utilizes NLP for the analysis of Windows logs, coupled with graph-based models that map out the interactions between different system events. This innovative approach is designed to reconstruct detailed attack narratives, providing a deeper understanding of cybersecurity threats.

Additionally, the research introduces enhancements to the machine learning model through stages such as graph construction, community detection, attack score calculation, topic modeling, and label propagation. These advancements are geared towards achieving a comprehensive analysis of cyber threats by examining the complex web of data relationships. By employing APT3 and classic playbooks to generate realistic simulated attack data, the study showcases Fuchikoma's proficiency in detecting malicious commands with remarkable precision. This is demonstrated by achieving over an 80% true positive rate and true negative rate, alongside an F3-score exceeding 60%.

UHAC (Kayhan et al., 2023): This study introduces the unsupervised hunting of anomalous commands (UHAC) system, which uses an unsupervised ML method designed to detect anomalous commands within SIEM logs. UHAC leverages the power of autoencoders trained on a unique combined feature set derived from both document-term and document-character matrices. By parsing text-based commands at two distinct levels – term and character — UHAC captures a comprehensive representation of command data, facilitating the detection of anomalies through the autoencoder's reconstruction capabilities. UHAC's methodology encompasses a preliminary filtering step for single-term commands, the creation of a combined feature set, and the development of an autoencoder-based detector utilizing a custom loss function tailored for anomaly detection. This unsupervised approach is particularly suited for the cybersecurity domain, where labeled data for training is scarce, and threats continuously evolve.

The effectiveness of UHAC was demonstrated through its application to real data from VirusTotal behavior reports, comprising over 50,000 entries. The system consistently outperformed other anomaly detection methods, including one-class SVMs, density-based spatial clustering (DBSCAN), and word embedding models like Word2Vec. In tests, UHAC successfully identified 84%–89% of anomalies within the top 10% of evaluated data, showcasing its robustness and efficiency in identifying potential threats within SIEM logs. UHAC's unsupervised learning model, combined with its innovative feature set, presents a significant advancement in the field of cybersecurity, offering a scalable and effective solution for enhancing threat detection mechanisms in SIEM systems. This work opens new avenues for research and application in anomaly-based cyber threat hunting, highlighting the potential of machine learning techniques in addressing the challenges of modern cybersecurity landscapes.

UN-AVOIDS (Yousef et al., 2021): This paper introduces the unsupervised and nonparametric approach for visualizing outliers and invariant detection scoring (UN-AVOIDS) framework, aimed at enhancing outlier detection through a unique integration of visualization and scoring mechanisms. UN-AVOIDS addresses the critical need for tools that can both detect and visualize anomalies in data, particularly in the realm of cybersecurity. The approach centers on transforming data into the neighborhood cumulative distribution function (NCDF) space, where outliers are visually distinguishable from the norm, facilitating both manual review and algorithmic analysis without the need for training or labeling. This transformation allows UN-AVOIDS to assign nuanced, continuous scores to data points based on their deviation in the NCDF space, providing a more flexible and informative metric for anomaly detection than traditional binary classifications.

Evaluated against established methods such as local outlier factor (LOF), isolation forest (IF), and fast angle-based outlier detection (FABOD), UN-AVOIDS demonstrated superior performance across a variety of scenarios. Its effectiveness was confirmed through tests on simulated data and two recent cybersecurity datasets, where UN-AVOIDS consistently outperformed comparative methods, as evidenced by higher AUC scores.

4.3. Reasoning techniques

Unlike traditional ML techniques that primarily focus on pattern recognition and anomaly detection based on historical data, reasoning techniques incorporate a more sophisticated approach, such as game theory and adversarial reasoning, to anticipate and counteract the strategies of attackers. The threat hunting literature presents several approaches to enhance detection and analysis capabilities. These methodologies leverage advanced technologies and reasoning techniques to generate or deduce the hypotheses in a contextual assessment. The reasoning techniques are classified as inductive reasoning (bottom-up approach) or deductive reasoning (top-down approach) (Mohamad et al., 2022). The details of the papers presented in this section are shown in Table 4.

CCS (Narayanan et al., 2018): The paper suggests a knowledge graph reasoner, using deductive reasoning techniques in combination with a knowledge graph to represent various steps of cyber attacks, linking these steps to related tools, techniques, and indicators detected by sensors.By associating specific tools such as Nmap with particular attack steps and observing their detection by systems like Snort, the theory-driven approach deduces potential attack stages. Deductive reasoning over the knowledge graph allows for identifying different steps in the attack chain, enhancing detection capabilities. This method increases confidence in identifying an attack as more and more indicators are linked, aids in assimilating information from diverse sources detects attack variations, and identifies new attacks that share characteristics with known ones. However, the current method is limited by its reactive nature, which means that a response is generated only after threats have occurred and relies on the record of past attack patterns. Therefore, the study proposes using the dataset collected from the dark net and hacker communities to expand the record of attack patterns for better response when threats are detected.

Socio-personal (Marin et al., 2020): The paper utilizes inductive reasoning techniques, such as causal reasoning and logic programming, to analyze hackers and their strategies. It captures socio-personal and technical indicators to generate corresponding rules based on this information. These rules are then used to identify potential cyberattacks. The proposed model shows a high F1 score (up to 150.24%) in comparison with the baseline predictor. The framework's deductive reasoning method is limited by its reliance on pre-established rules and patterns, potentially overlooking new or evolving attack strategies. Future work aims to integrate adaptive and learning-based approaches to improve predictive accuracy and adapt to emerging cyber threats.

Game theory (Dritsoula et al., 2017): The paper demonstrates game theory models in cybersecurity to analyze the interactions between attackers and defenders as a strategic game, allowing for the prediction of potential attacks and the formulation of optimal defense strategies. The game-theoretic model's primary limitation is its assumption of rational behavior from attackers and defenders. Future work should consider scenarios where attackers may behave irrationally or unpredictably, enhancing the model's robustness and applicability.

4.4. Graph-based approaches

This section presents an overview of various advanced graph-based methodologies and cyber threat detection and analysis frameworks. Various types of graphs are utilized, including Query Graphs, Provenance Graphs, Behavior Provenance Graphs (BPGs), Knowledge Graphs, Heterogeneous Information Networks (HIN), Semantic Association Graphs, and Event Graphs. These approaches leverage graph theory and machine learning to model complex relationships between cyber entities (e.g., indicators of compromise (IOCs), system events, threat actors). They integrate threat intelligence, visualizing attack progressions, and extracting key events to identify and analyze sophisticated cyber threats. The details of the papers presented in this section are shown in Table 5.

Poirot (Milajerdi et al., 2019a): This study introduces a system for enhancing cyber threat hunting through the innovative use of CTI and kernel audit records. Poirot distinguishes itself by effectively leveraging the relationships between IOCs, which are often neglected in traditional threat hunting methodologies. The core approach involves constructing a *query graph* from CTI, outlining the expected behavior of an attack, and a *provenance graph* from kernel audit logs, detailing the actual system activities.

The essence of Poirot's methodology is a graph alignment technique, where the system aligns the query graph with the provenance graph to identify potential threats. This alignment is facilitated by a unique similarity metric that evaluates the correspondence between the two graphs, enabling Poirot to pinpoint attack activities with high accuracy within large-scale datasets.

Poirot demonstrates its capability to search through extensive graphs efficiently, identifying attack behaviors within minutes. Tested on a variety of datasets, including real-world incident reports and controlled adversarial engagements, Poirot showed remarkable effectiveness in detecting nuanced attack patterns, underscoring the utility of CTI correlations as robust artifacts for threat hunting.

DeepHunter (Wei et al., 2021): This paper introduces a method to enhance cyber threat detection through the application of graph neural networks (GNNs). The study addresses the challenge of inconsistencies between actual attack activities and known attack behaviors within provenance data, a common obstacle in cyber threat hunting. DeepHunter utilizes a GNN model for graph pattern matching, effectively estimating the matching score between a provenance graph and a query graph. The approach is innovative, incorporating attribute embedding networks to integrate information about IOCs, and graph embedding networks to capture the intricate relationships between IOCs, thereby providing a robust solution to the inconsistency problem.

The evaluation of DeepHunter across five APT attack scenarios, including both real and synthetic datasets, demonstrates its ability to identify all attack behaviors (claimed). This effectiveness is notably higher than that of Poirot (Milajerdi et al., 2019a).

LogKernel (Li et al., 2022b): This study introduces a cyber threat hunting method that identifies attack behaviors in system audit logs without relying on CTI. LogKernel abstracts system activities into behavior provenance graphs (BPGs) and employs a custom graph kernel clustering technique to analyze these graphs. By capturing both the structural information and label data of BPGs, LogKernel effectively quantifies the similarity between behaviors, facilitating accurate clustering and identification of potential threats.

The method was evaluated using a malicious dataset containing simulated attack scenarios and the DARPA CADETS dataset, which includes real attack instances. LogKernel demonstrated capability in accurately detecting both known and unknown attacks, successfully identifying all attack scenarios within the datasets. LogKernel's approach focuses on structural analysis and similarity-based clustering of behavior provenance graphs.

Extractor (Satvat et al., 2021): This study introduces an innovative system designed to transform unstructured CTI reports into structured, concise representations of attack behaviors using provenance graphs. This transformation is crucial for enhancing the utility of CTI in cybersecurity operations, particularly in threat hunting and incident response. Extractor tackles the challenges of verbosity, text complexity, and relationship extraction inherent in processing CTI reports. The system applies a combination of text summarization techniques, semantic role labeling (SRL), and graph generation methods to distill relevant attack information, identify key actions and entities, and construct provenance graphs that accurately represent the sequence and dependencies of attack behaviors. The evaluation of Extractor utilized real-world incident reports and DARPA adversarial engagement reports.

ANUBIS (Anjum et al., 2022): This paper presents a framework designed to enhance the detection of APTs through the innovative use of system provenance graphs and Bayesian neural networks (BNNs). By constructing detailed provenance graphs that capture the causal relationships between system events, ANUBIS effectively identifies sequences of activities indicative of APTs, offering a nuanced understanding of malicious operations within a system. The approach integrates machine learning into the cyber threat detection process, specifically

Table 5
Graph-based approaches in threat hunting.

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
Milajerdi et al. (2019a)	Graph pattern matching	Publicly released real-world incident reports and DARPA adversarial engagement scenarios	Uses CTI correlations and kernel audits to detect attack patterns	Runtime increases as the size of the query grows
Wei et al. (2021)	Graph Neural Networks (GNNs)	Five APT attack scenarios with both real and synthetic datasets	Addresses the challenge of inconsistencies between actual attack activities and known attack behaviors within provenance data	–
Li et al. (2022b)	Behavior Provenance Graphs (BPGs)	Simulated attack scenarios and the DARPA CADETS dataset	Identifies known and unknown attack scenarios within the datasets	Its dependency on trusted audit logs, inability to detect non-system call interface attacks and OS kernel vulnerability exploits, manual involvement in label definition, and less accurate density-based partitioning. Future work involves addressing data imbalance in threat hunting and scaling to other logging formats.
Gao et al. (2022)	Meta-Path and Meta-Graph Instances-Based Similarity Measure (MIIS) and Graph Convolutional Network (GCN)	IBM X-Force Exchange (Brown et al., 2015) and VirusTotal	Automates CTI analysis and reduce the manual workload on security analysts	It considers only a limited number of infrastructure node types and relations and disregard the dynamic threat type changes. Future work involves enriching node features and relations and extracting fine-grained structured data from natural language intelligence reports using NLP and topic modeling.
Satvat et al. (2021)	Provenance Graphs and semantic role labeling (SRL)	Real-world incident reports and DARPA adversarial engagement reports	Identifies key actions and entities, and construct provenance graphs that accurately represent the sequence and dependencies of attack behaviors	The limitations include challenges with action descriptions spanning multiple sentences, inability to detect timing and side-channel attacks, reliance on dictionaries that may be incomplete, and its focus on natural language descriptions. Future work involves improving dictionaries, incorporating Named Entity Recognition, and extending the model to infer graphs from unstructured vulnerability reports for enhanced vulnerability detection.
Anjum et al. (2022)	Provenance graphs and Bayesian Neural Networks (BNNs)	DARPA OpTC dataset, a comprehensive collection of APT-like activities	Detects sophisticated cyber threats with minimal computational overhead	Dependence on data from a single IT infrastructure, need for large training data volumes, sensitivity to data poisoning attacks, and a focus on host-based analysis. Future work involves testing on diverse technological stacks, improving training processes, ensuring data integrity, and exploring universal provenance graphs for comprehensive APT detection.
King and Huang (2023)	GNNs and RRNs	DARPA OpTC dataset, which simulates APT-like activities	Predicts potential unauthorized lateral movements by analyzing the causal relationships between network activities	Its dependency on sufficient benign data for baseline learning, potential contamination of training data with malicious activity, difficulty in determining the appropriate granularity of graph snapshots, and potential delays in detection time. Future work involves exploring the use of a greedy interval partition algorithm for job assignments, enhancing feature extraction for system entities, and determining the optimal granularity for graph time slices to balance informativeness and training efficiency.
Kaiser et al. (2023)	Knowledge graph traversal and link prediction techniques	Over 53,000 VirusTotal reports	Refines and generates hypotheses about potential attacks based on observable network and system artifacts	Future work involves empirically evaluating the presented algorithms against professional security analysts and rule-based TTP inference like HOLMES (Milajerdi et al., 2019b), expanding rule bases for comprehensive ATT&CK technique coverage, incorporating sequence information from CTI sources into AttackDB, and developing automation to close the loop with data collection based on high-level attack hypotheses.

(*continued on next page*)

Table 5 (*continued*).

Study	Methods	Datasets	Approach uniqueness	Limitations and future works
Wang et al. (2022a)	Provenance graphs and GraphSAGE	DARPA Transparent Computing program, the CAIDA DDoS Attack 2007 dataset, the CICIDS2017 dataset by the Canadian Institute for Cybersecurity, and the ADFA-LD dataset from the Australian Defense Force Academy	Detects abnormalities within benign system operations	Closed-world assumption (Sommer and Paxson, 2010), vulnerability to various adaptive attacks, reliance on potentially unrepresentative datasets, system overhead management, threat fatigue from false positives, and the need for benign data in its semi-supervised approach. Future work focuses on addressing these issues through improved model updates, robustness against diverse attacks, better dataset generation, system performance adjustments, reduction of false positives, and exploring unsupervised learning methods.
Li et al. (2023b)	Provenance graphs	DARPA datasets and simulated practical APT attack scenarios	Develops a training-free solution that focuses on the extraction of attacker activities directly from raw logs	The need for further testing across different platforms and systems and the exploration of network event identification from raw logs. Future work will focus on expanding the approach's application, enhancing attack prevention guidance, and conducting more experiments with diverse datasets.
Zang et al. (2023b)	Semantic association graphs and community detection algorithms	DARPA 2000, CICIDS 2017, and real internet traffic from the China Education Research Network backbone (CERNET)	Visualizes the progression of multi-stage attacks and identifies related groups of threat indicators that constitute an attack scenario	–
Ho et al. (2021)	Graph inference algorithm	Commonly available enterprise logs (780 million internal logins within a 15-month period at a large enterprise)	Introduces a specification-based anomaly detection approach	Challenges with detecting stealthy attacks that exploit legitimate user logins and inaccuracies in logging data. Future work aims to address these issues by exploring evasion strategies, improving generalization across different network architectures, enhancing detection performance, and incorporating additional data sources and commercial log-hygiene solutions.

employing a BNN to analyze encoded event traces derived from the provenance graphs. This method allows ANUBIS to classify these traces as benign or malicious with a high degree of accuracy, while also providing confidence levels for its predictions to aid explainability.

Evaluated on the DARPA OpTC dataset, a comprehensive collection of APT-like activities, ANUBIS showcased exceptional performance, achieving an accuracy rate of 99% and a precision over 98%, with a false positive rate of less than 2%. These results underline the system's ability to efficiently and reliably detect sophisticated cyber threats with minimal computational overhead.

The core contribution of ANUBIS lies in its combination of provenance graphs with advanced machine learning techniques, offering a sophisticated and highly effective tool for APT detection. The use of a BNN, in particular, enhances the framework's predictive power and introduces an element of explainability into the threat detection process, making it an invaluable asset for cyber-response teams.

Euler (King and Huang, 2023): This study introduces an approach to detecting lateral movement in network systems, a key strategy used in APTs. The proposed framework, named Euler, applies temporal link prediction techniques using GNNs combined with sequence encoding layers such as RNNs to model and analyze the dynamic behavior of network interactions over time. This methodology allows for the effective identification of anomalous activities that may signify lateral movements by attackers within a network. The core of Euler's approach lies in its ability to process and analyze system provenance graphs, which detail the causal relationships between network activities, to predict potential unauthorized lateral movements. By framing the detection of such movements as a problem of identifying anomalous links within these evolving graphs, Euler can pinpoint unusual patterns indicative of cyber threats with high accuracy.

Evaluated on the DARPA OpTC dataset, which simulates APT-like activities, Euler demonstrated exceptional performance, achieving a 99% accuracy rate and over 98% precision, while maintaining a false positive rate of less than 2%. These results significantly surpass those of existing unsupervised techniques, highlighting Euler's effectiveness and efficiency in lateral movement detection.

AttackDB (Kaiser et al., 2023): This paper introduces a novel framework aimed at automating the generation of attack hypotheses in cybersecurity threat intelligence analysis. The framework is comprised of the AttackDB knowledge base and the attack hypothesis generator (AHG), and leverages comprehensive threat intelligence from multiple open-source datasets, including MITRE ATT&CK Enterprise, AlienVault Open Threat Exchange, IBM X-Force Exchange, and VirusTotal, to create a detailed knowledge graph. This graph maps high-level ATT&CK techniques to low-level observable artifacts found in behavioral malware reports, enabling the automated inference of adversarial techniques from vast amounts of data.

Rather than utilizing conventional ML models, the approach employs knowledge graph traversal and link prediction techniques to refine and generate hypotheses about potential attacks based on observable network and system artifacts. This methodology allows for the systematic and automated identification of adversarial techniques, significantly aiding analysts in the threat detection and analysis process.

Experimental results, based on over 53,000 VirusTotal reports, demonstrate the effectiveness of the AHG in accurately inferring adversarial techniques, achieving a mean average precision greater than 50% and an AUC of over 0.8.

THREATRACE (Wang et al., 2022a): This study introduces an advanced system for identifying and tracing stealthy host-based cyber threats, including program attacks, malware implantation, and APTs, at the individual system entity level. This approach leverages the power of data provenance graphs — directed acyclic graphs that detail system entities and their interactions — combined with GraphSAGE, a graph neural network designed for inductive learning.

Unlike traditional methods that rely on known attack patterns, threaTrace focuses on detecting abnormalities in benign system operations, enabling the early detection of sophisticated cyber threats that subtly alter system behavior. The methodology centers on utilizing

GraphSAGE in provenance graphs to accurately model and understand normal system entity behavior. This allows threaTrace to identify deviations indicative of malicious activity, addressing the challenges of data imbalance and the need for early intrusion detection.

The THREATRACE system was evaluated across five public datasets encompassing a broad spectrum of APT-like activities such as the DARPA Transparent Computing program, the CAIDA DDoS Attack 2007 dataset, the CICIDS2017 dataset by the Canadian Institute for Cybersecurity, and the ADFA-LD dataset from the Australian Defense Force Academy for Linux-based intrusion detection, among others.

THREATRACE consistently demonstrates strong performance across various DARPA TC datasets. Specifically, in the scene #3/THEIA, it achieved precision and recall rates of 0.87 and 0.99, respectively, culminating in an F1 score of 0.93. Additionally, the framework reached a precision score of 0.90 in the scene #3/CADETS, maintaining a high recall of 0.99. However, despite these successes, THREATRACE exhibits some variability in precision on some other datasets, with the lowest precision of 0.63 reported in the scene #5/CADETS dataset alongside a recall of 0.86. This variability indicates potential challenges in consistently minimizing false positives across varied datasets.

T-trace (Li et al., 2023b): This paper introduces a groundbreaking approach for tracing APTs through the construction of provenance graphs by correlating multiple system logs (syslogs). This method addresses critical challenges in cybersecurity threat analysis, including the need for efficient pattern learning, semantic correlation, and overcoming the dependency explosion problem that plagues conventional log analysis techniques. T-trace employs tensor decomposition to mine the implicit relationships in vast quantities of log data, identifying log communities associated with specific log templates. This process, coupled with the calculation of significance scores, allows for the extraction of key events and the discovery of event communities through log correlation.

Unlike traditional machine learning-based methods for tracing APT activities, T-trace operates without the need for a predefined training dataset, offering a training-free solution that focuses on the extraction of attacker activities directly from raw logs. The framework was rigorously tested using DARPA datasets and simulated practical APT attack scenarios, demonstrating T-trace's ability to significantly reduce analysis time by 90% while achieving an accuracy rate of 92% in the construction of provenance graphs. These results highlight T-trace's efficiency and precision in identifying specific APT-related events and constructing detailed graphs that map attacker movements within a system.

Multi-step attack (Zang et al., 2023b): This study introduces a framework for reconstructing multi-step attack scenarios by integrating diverse sources of threat intelligence. The method aims to address the complexities involved in understanding cyber attacks that unfold in stages, which are often difficult to trace due to the disparate nature of the data collected by various security systems. By leveraging the STIX format, the framework standardizes and fuses threat intelligence, enabling the analysis of causal relationships among different threat indicators. The approach involves constructing semantic association graphs that map out the connections between various pieces of threat intelligence, effectively visualizing the progression of multi-stage attacks. Further, the method treats attack scenario reconstruction as a community discovery problem, applying community detection algorithms to identify related groups of threat indicators that constitute an attack scenario. The evaluation of this framework utilized open-source benchmark datasets (DARPA 2000, CICIDS 2017) (Sharafaldin et al., 2019) and real internet traffic from the China Education Research Network backbone (CERNET) (Zang et al., 2023a).

Hopper (Ho et al., 2021): This paper presents an innovative approach to detecting lateral movement within enterprise networks, a critical phase in sophisticated cyber attacks where attackers navigate through a network to escalate privileges or reach valuable targets. Hopper leverages commonly available enterprise logs to construct a graph of login activities, identifying suspicious sequences that signify lateral movements. Unlike traditional methods that focus on anomaly detection or rely on narrowly crafted signatures, Hopper introduces a specification-based anomaly detection approach. It identifies fundamental characteristics of lateral movement, focusing on sequences that involve credential switching and access to servers beyond the attacker's initial compromise. Hopper employs a graph inference algorithm to discern the broader paths of user movements, identifying the causal user behind each path. By analyzing these paths in the context of specified attack properties — such as credential switching and accessing previously inaccessible servers — Hopper effectively pinpoints activities indicative of lateral movement.

The system's efficacy was evaluated on a dataset comprising over 780 million internal logins in 15 months at a large enterprise, achieving a 94.5% detection rate across more than 300 realistic attack scenarios, including a professional red team attack, while maintaining a low false positive rate.

HinCTI (Gao et al., 2022): This paper introduces a groundbreaking system that leverages CTI to enhance cybersecurity defenses. HinCTI is designed around the concept of a heterogeneous information network (HIN) to model CTI, incorporating a variety of cyber-threat infrastructure nodes and their intricate relationships. This approach addresses the challenges posed by the complexity and heterogeneity of cyber-attacks and their underlying infrastructure. The methodology involves three key components: (i) employing HIN for CTI modeling to capture both explicit and implicit relationships among diverse types of infrastructure nodes; (ii) utilizing meta-path and meta-graph instances-based similarity measure (MIIS) for accurately measuring the similarity between threat infrastructure nodes; and (iii) applying a heterogeneous graph convolutional network (GCN) for the identification of threat types, enhanced with a hierarchical regularization strategy to mitigate overfitting and improve model performance.

HinCTI's efficacy was demonstrated through extensive experiments on real-world data collected from platforms such as IBM X-Force Exchange (Brown et al., 2015) and VirusTotal (Salem et al., 2021). The system significantly outperformed existing methods in threat type identification, showcasing its potential to automate CTI analysis and reduce the manual workload on security analysts.

4.5. Rule-based approaches

This section explores rule-based approaches aimed at bolstering cyber threat detection and analysis within enterprise networks. These methodologies leverage predefined rules and patterns derived from threat intelligence to identify known adversarial tactics, techniques, and procedures (TTPs) evident in network traffic and system behavior. The techniques used include rule-based matching, anomaly detection, provenance graph construction, measurement of behavior deviations, multi-level provenance modeling, and community discovery. The aim is to utilize predefined rules and patterns from threat intelligence to detect, correlate, and interpret complex attack patterns across network activities and system logs.

SteinerLog (Bhattarai and Huang, 2022): This study presents a novel system designed to enhance cyber threat detection in enterprise networks by automating the correlation of alerts to uncover APT campaigns. Leveraging an approach that combines rule-based matching, anomaly detection, and provenance graph construction, SteinerLog facilitates a hierarchical analysis of both intra-host and inter-host activities to reconstruct and understand attacker movements across the network.

Rule-based matching allows for the identification of suspicious events by comparing system activities against a knowledge base of known adversarial TTPs, sourced from community-driven threat intelligence repositories. Anomaly detection complements this by flagging deviations from established patterns of normal behavior, capturing potentially malicious activities not documented in existing TTPs. The core of SteinerLog's methodology involves constructing provenance graphs for each host, depicting the system execution history to provide rich contextual information for detected alerts. Through hierarchical correlation and analysis, SteinerLog correlates attacker activities across the network, identifying compromised entities and abstracting complex attack behaviors into intuitive attack graphs. SteinerLog demonstrates potential in generating and scoring online attack campaigns for APT detection through the evaluation of data from a simulated enterprise network environment utilizing Operational Transparent Cyber (OpTC). Nonetheless, it encounters hurdles concerning detection accuracy and limitations within simulations. The system relies on detectors to capture attacker activities, which might be evaded by well-resourced adversaries. Anomaly detectors, while filling gaps left by rule-based systems, do not tag anomalies with TTPs, requiring manual inspection and depending heavily on training data quality. Simulated attack campaigns, conducted over hours by security experts, do not fully mimic the prolonged, stealthier real-world APT campaigns, affecting the system's ability to correlate events over extended periods and identify subtle evasion efforts.

To address these challenges, modifications are suggested: implementing a hierarchical storage and caching mechanism to handle extensive provenance graphs and utilizing techniques to prune causally irrelevant data from these graphs. These adjustments aim to enhance SteinerLog's ability to detect prolonged, low-intensity attacks without compromising system performance or risk-scoring accuracy.

C-BEDIM/S-BEDIM (Dong et al., 2023): This paper presents a pioneering approach to detecting lateral movement within enterprise networks, a critical aspect of identifying sophisticated cyber threats. This work introduces two methodologies, simple behavior deviation measurement (S-BEDIM) and complex behavior deviation measurement (C-BEDIM), which leverage behavior deviation measurements to identify unusual network activities potentially indicative of an attacker's lateral movements.

The methodologies commence by constructing graph sequences from network access records, converting these into a graphical format that encapsulates the network's connection dynamics over time. By employing connection expansion, the approaches broaden the analysis scope to include second-order connections, enhancing the detection capabilities. Deviation scores are subsequently calculated for each node to quantify behavioral abnormalities, with high-scoring nodes undergoing further analysis to generate high-quality alerts indicative of potential lateral movement.

The evaluation of real-world datasets from the intranets of two enterprises demonstrated the effectiveness of S-BEDIM and C-BEDIM in accurately identifying lateral movement activities. Notably, C-BEDIM achieved a 100% detection accuracy under specific conditions, outperforming traditional detection methods and underscoring its potential as a robust solution for enterprise security.

ProvTalk (Tabiban et al., 2022): This study presents a novel system designed to enhance the security incident analysis in complex networking functions virtualization (NFV) environments. ProvTalk introduces a multi-level provenance model that effectively captures the intricate dependencies across various NFV layers, enabling a comprehensive understanding of security incidents by constructing detailed provenance graphs from system logs.

The approach adopted by ProvTalk is multifaceted, comprising the following stages:

Multi-level Provenance Construction: A method to encapsulate the dependencies between different NFV levels into a coherent provenance graph, providing a holistic view of system activities.

Multi-level Pruning: A technique to improve the interpretability of provenance graphs by pruning irrelevant information, focusing analysis on the nodes critical for understanding the security incident.

Mining-based Aggregation: An innovative aggregation method that condenses redundant information within the provenance graphs, simplifying the analysis by highlighting significant patterns and behaviors.

Rule-based Natural Language Translation: A strategy to translate the technical details encapsulated in the provenance graphs into human-readable text, facilitating easier interpretation and analysis by security analysts.

ProvTalk was evaluated in a real-world scenario using the Tacker-OpenStack NFV platform, where it demonstrated significant capabilities in reducing the complexity of provenance graphs while retaining essential information for security analysis. The system achieved approximately a 3.6 times reduction in graph size through multi-level pruning and a two times reduction via aggregation. This study leverages NLP techniques and rule-based approaches for the innovative translation of provenance graphs into interpretable formats.

HERCULE (Berady et al., 2021): This paper demonstrates a framework for analyzing system logs to reconstruct narratives of multi-stage cyber attacks. The HERCULE system leverages community discovery techniques within a graph analytics framework to correlate disparate log entries across various system logs. By modeling multi-stage intrusion analysis as a community discovery problem, HERCULE identifies "attack communities" that represent sequences of related attack steps within the generated multi-dimensional weighted graphs. The approach begins with the construction of causal dependency graphs from system logs, aiming to capture the intricate relationships and sequences of events indicative of cyber attacks. Utilizing advanced graph analytics and community detection algorithms, HERCULE efficiently segregates these sequences into coherent narratives that detail the progression of attacks across multiple stages. Evaluated against a dataset comprising well-known APT attack families, HERCULE demonstrated its effectiveness in accurately identifying and reconstructing attack narratives, showcasing high precision, recall, and F1 scores. This success underscores the framework's capability to discern between benign and malicious activities within extensive system logs, significantly reducing false positive rates.

4.6. Other methods

Other threat hunting methodologies utilize miscellaneous techniques including behavioral analytics, reinforcement learning (RL), and various statistical methods.

UEBA (Shashanka et al., 2016): This study presents an approach to enhancing enterprise cybersecurity through the deployment of user and entity behavior analytics (UEBA). By leveraging advanced machine learning techniques, specifically singular value decomposition (SVD) and Mahalanobis distance, this intelligence platform is designed to monitor and analyze the behaviors of users, IP addresses, and devices within an enterprise network to identify potentially malicious activities. This methodology enables the detection of anomalies by comparing observed behaviors against established baselines, flagging deviations for further investigation. The system's effectiveness is demonstrated through the application of SVD for data transformation and dimensionality reduction, coupled with the use of Mahalanobis distance for quantifying the extent of deviation from normal behavior patterns.

While the paper details the theoretical underpinnings and operational mechanisms of these techniques, it does not provide specific quantitative results from the deployment of the UEBA modules. Instead, it offers empirical examples to illustrate the system's capability in detecting anomalous activities, underscoring the potential for the early identification and mitigation of security threats. Although explicit datasets are not mentioned, the application of these machine learning algorithms implies the analysis of comprehensive data sources within the enterprise, such as network logs and packets, to model normal behavioral patterns and identify outliers.

CTI Ontology (Mavroeidis and Jøsang, 2018): This paper introduces an automated threat assessment system designed to enhance cybersecurity defenses by analyzing continuous incoming feeds of Sysmon logs. By leveraging the detailed visibility provided by Sysmon logs, the system classifies software into various threat levels in real-time, thus enabling organizations to identify and respond to potential cyber threats more effectively. At the heart of the system is a CTI ontology (CTIO), which integrates and represents diverse sources of threat intelligence, thereby supporting informed decision-making processes. The approach is grounded in a review of current threat intelligence practices and the development of an agile, ontology-based framework that encapsulates a wide array of cyber threat data — from low-level technical observables to high-level actor strategies and goals. The system's performance and operational flow are demonstrated through practical applications, highlighting its capability for situational awareness, threat prediction, and the execution of automated courses of action.

MABAT (Dekel et al., 2023): This paper introduces the multi-armed bandit approach for threat-hunting (MABAT) framework, which leverages multi-armed bandit (MAB) techniques to optimize the process of targeted data collection for cyber threat hunting. This approach is predicated on the need to efficiently gather and analyze data related to potential cyber threats, balancing between exploring new data and exploiting known information to maximize the relevance of collected artifacts. MABAT introduces an application of combinatorial multi-armed bandit problems, specifically the multi-shared-arms bandits variant, to the domain of cyber threat hunting. By employing augmented MAB policies that consider the shared attributes among various attacks, MABAT enhances the ability to identify and focus on the most pertinent attack vectors based on CTI.

The effectiveness of MABAT was demonstrated through an evaluation utilizing real behavioral reports from VirusTotal, containing inconsistencies and varied degrees of attack representation. This technique can be classified under the umbrella of reinforcement learning due to its focus on optimizing a decision-making process through learning.

ELK stack (Almohannadi et al., 2018): This study proposes an approach for generating CTI by deploying honeypots to collect data on potential cyber attacks. The approach utilizes the ELK stack (Elasticsearch, Logstash, and Kibana) to analyze and visualize the data obtained from honeypots, aiming to uncover attack patterns and behaviors. By analyzing honeypot log data, the study seeks to enhance the understanding of cyber threats and improve the intelligence available for cybersecurity defenses. The methodology centers around the collection of cyber incident log data from honeypots deployed in an AWS cloud environment. The log data, which encompasses a variety of cyber attack attempts, is then processed and analyzed using elastic search technology. The use of the ELK stack enables efficient data analysis and visualization, facilitating the identification of malicious activities within the voluminous data generated by honeypots. While the research outlines the theoretical framework and potential of using honeypots and elastic search for threat intelligence, it does not provide specific empirical results or performance metrics from the analysis. Rather, the focus is on demonstrating the feasibility and conceptual benefits of the proposed approach for enhancing CTI through the innovative use of technology.

Hidden communication channels using steganography: Operation Stegoloader is a notable example of cybercriminals using steganography for data exfiltration and malware concealment. Stegoloader, also known as "TSPY_Gatak", is a type of malware that uses steganographic techniques to evade detection, first revealed in 2015 (Mimoso, 2015).

However, there are a few campaigns, such as UAC-0184, that are actively using steganography techniques to embed malicious code within image files (IDAT loader 2024). The TA558 threat actor conducted the SteganoAmor campaign, targeting over 320 organizations globally, with a focus on Latin America. This campaign used steganography to embed base64-encoded payloads within JPG image files. These images were delivered through phishing emails exploiting an old Microsoft Office vulnerability (CVE-2017-11882). The payloads included various malware families such as AgentTesla, FormBook, Remcos, LokiBot, Guloader, Snake Keylogger, and XWorm. This wide-ranging campaign exploited compromised SMTP servers and legitimate cloud services to distribute and control the malware (Technologies, 2024; Abrams, 2024). Several papers have extensively explored the topic of hidden data communication, focusing on the analysis of various neural network architectures to identify potential hidden communication channels (Dzhanashia and Evsutin, 2024; Lerner and Romanov, 2022; Melman and Evsutin, 2024).

5. Challenges

The objective of this section is to tackle research question **RQ4** by elucidating the challenges inherent in existing threat hunting models as documented in the literature. Despite substantial advances in the threat hunting field over recent decades, practitioners continue to grapple with numerous challenges that stymie their efforts towards more effective detection and prevention of malicious activities. These challenges not only underscore the complexity and evolving nature of cyber threats but also highlight the gaps in current methodologies, technologies, and analyst skills. This section summarizes the primary obstacles encountered in threat hunting, shedding light on the intricacies of navigating a landscape where adversaries constantly refine their strategies to elude detection. As such, the summary provides a foundational understanding of the hurdles that must be overcome to enhance the efficacy and efficiency of threat hunting practices. The major challenges encountered in the context of threat hunting include, but are not limited to:

1. a lack of labeled data;
2. imbalanced datasets;
3. multiple sources of log data;
4. adversarial techniques;
5. a scarcity of human experts and data intelligence.

5.1. Challenge 1: A lack of labeled data makes it difficult to train trustworthy threat detection models

Anomaly detection has long been framed as a classification problem in event logs, distinguishing normal activities from abnormal ones (Agarwal et al., 2021; Apruzzese et al., 2023a). Various approaches have been employed, including supervised and unsupervised models, spanning traditional ML to deep learning (DL). Supervised learning techniques have been widely developed, involving training with labeled data to classify event logs and identify malicious activities. These techniques encompass a range of methodologies, from classical shallow models such as decision trees (Khraisat et al., 2020) and support vector machines (Mukkamala et al., 2002) to more complex deep learning architectures like convolutional neural networks (CNNs) (Wang et al., 2020; Vinayakumar et al., 2017) and recurrent neural networks (RNNs) (Yin et al., 2017; Woźniak et al., 2020). Using labeled data, supervised learning enables the construction of highly accurate models capable of effectively detecting a wide range of malicious activities. However, relying on labeled data presents its own challenges. It is time-consuming and resource-intensive to acquire and annotate large volumes of labeled data, especially since cyber threats are dynamic and constantly evolving. Furthermore, the availability of labeled data may be limited, especially for emerging or rare threat scenarios, thereby offsetting the effectiveness of supervised learning strategies.

In addressing these challenges, novel techniques such as self-supervised, online learning and federated learning have been recently introduced. For example, Self-Supervised Intrusion Detection (SSID) (Nakıp and Gelenbe, 2024) is a recently proposed framework, which enables a fully online Deep Learning (DL) based Intrusion Detection System (IDS) that requires no human input or prior offline training. Utilizing an Auto-Associative Deep Random Neural Networkmodel (Nakip and Gelenbe, 2021; Gelenbe and Nakıp, 2022), the proposed framework classifies and labels incoming traffic packets based on the IDS's decisions. With this approach, the IDS can continuously adapt to new network conditions and threats, ensuring its performance without pre-labeled data. For botnet attack detection in IoT devices, Shao et al. (2021) proposes an adaptive online learning strategy that enables the detection model to adapt to pattern changes in IoT traffic in real-time. This approach addresses concept drift and dynamic traffic patterns, improves detection performance by employing ensemble learning, and reduces reliance on pre-labeled training data. Abououf et al. (2022) proposes a lightweight anomaly detection system for IoT devices uses a Long Short-Term Memory Autoencoder (LSTM AE). The approach reduces the reliance on extensive communication with the server and utilizes smart inference techniques to detect anomalies without requiring large amounts of labeled data. Nakip et al. (2023) proposes the Decentralized and Online Federated Learning Intrusion Detection (DOF-ID) architecture, which enhances detection performance by leveraging collaborative learning across distributed system components. DOF-ID improves intrusion detection without reliance on extensive labeled data by allowing components to learn from experiences gained by other components, along with their own local data.

5.2. Challenge 2: Imbalanced datasets make learning generalization difficult

Imbalanced datasets pose another significant obstacle to threat detection. Imbalanced datasets arise when the distribution of labeled classes is highly skewed. Hence, this issue is closely linked to the need for labeled data in supervised learning techniques. The challenge is particularly prevalent in the cyber threat arena. In many real-world scenarios, instances of anomalous activities are relatively rare compared to normal activities. For example, every system that logs users and their daily activities generates a large number of events. However, while the majority of system events will represent legitimate user behavior, only a small fraction may indicate potential security threats (Zhang et al., 2022).

Despite the myriad approaches available, addressing imbalanced datasets in threat detection remains a daunting task (Chen et al., 2022b). The inherent challenge stems from the fact that any ML algorithm, be it supervised or unsupervised, is susceptible to the adverse effects of class imbalance. When trained on such datasets, models tend to exhibit biases towards the majority class, resulting in suboptimal generalization and an elevated risk of false negatives. Consequently, the ability to detect rare or emerging threats becomes markedly compromised. Efforts to mitigate this issue include either rebalancing the dataset (Moti et al., 2020) or reposting the problem as anomaly detection (Du et al., 2017b; Villarreal-Vasquez et al., 2021); however, the persistent nature of class imbalance underscores the complexity of the challenge and the need for continued innovation in the field of threat detection. In addition, the threat hunting datasets are susceptible to the temporal sensitivity concept where attack logs predominantly appear towards the end of the investigation period, highlighting a nuanced challenge in threat hunting and detection. Imbalanced datasets can pose significant challenges to threat hunting and detection in various scenarios, as outlined below:

Rare Event Detection: Cyber threats, by nature, are less frequent than normal network events. This rarity means that models trained on such data may have difficulty learning the characteristics of these rare events, leading to a high rate of false negatives, where actual threats go undetected.

Dynamic Nature of Threats: Cyber threats are not static; attackers continually develop new techniques to evade detection. As a result, the characteristics of malicious activities can change over time, leading to concept drift. This dynamic nature further complicates model training on imbalanced datasets, as models need to generalize well to previously unseen threats while dealing with an imbalance.

Cost of Misclassification: In cybersecurity, the cost of misclassifying a malicious event as benign (false negative) is often much higher than misclassifying a benign event as malicious (false positive). False negatives can allow attackers to continue their activities undetected, while false positives, although inconvenient, typically result in additional investigation. This asymmetric cost reinforces the need for models that are sensitive to the minority class in imbalanced datasets.

The literature suggested several techniques to overcome imbalanced data. Techniques such as the synthetic minority over-sampling technique (SMOTE) Chawla et al. (2002) support vector machine (SVM) technology (Akbani et al., 2004), cost-sensitive learning (Fernández et al., 2018), and K-nearest neighbor (KNN) technology each offer unique mechanisms for rebalancing data (He and Garcia, 2009). SMOTE, for instance, synthesizes new minority class instances by interpolating between existing ones, thereby enriching the dataset without losing valuable information. On the other hand, SVM technology and KNN technology apply modifications at the algorithmic level, altering how the classifier interprets the class distribution, thus making these models more robust to imbalances. Fig. 5 presents a generic diagram that abstracts the techniques used to deal with imbalanced data.

Beyond data-level interventions, algorithm-level methods provide a sophisticated framework for adapting the learning process itself to be more attuned to class imbalance. Single-class learning (Tax and Duin, 2004) and ensemble learning emerge as pivotal approaches in this category. Single-class learning focuses on the properties of one class, often the minority, to enhance its detectability, a technique particularly beneficial in scenarios where the minority class is of paramount importance. Ensemble learning, incorporating strategies based on both bagging and boosting, leverages the collective strength of multiple models to achieve a more balanced and accurate classification performance. Bagging methods like random forests reduce variance by training numerous decision trees on varied subsets of the data while boosting sequentially refines the model focusing on previously misclassified instances, thereby incrementally improving classification accuracy. These algorithm-level methods underscore the evolving complexity and sophistication in addressing imbalanced datasets, emphasizing a move towards more nuanced and adaptable machine learning solutions (Wang et al., 2021).

5.3. Challenge 3: Leveraging multiple sources of data enhances the effectiveness of threat detection

In cybersecurity, relevant data can originate from various sources, such as Windows event logs, Sysmon logs, and memory logs. Moreover, data streams from user machines, network infrastructures, and sentinel systems contribute to the rich tapestry of information available for analysis. Integrating and analyzing these diverse data sources cohesively is crucial to improving the effectiveness of threat detection. As different sources of data complement each other by providing both compatible and distinct types of information, integrating multiple modes of data in the learning model will boost the learning process to learn more comprehensive features or indicators to detect malicious event logs.

Various multi-source approaches have been developed for the cyber threat detection problem, for example, to enhance an IDS (Lin et al., 2022), or for mining association rules from snort logs, firewall logs, and system logs (Lou et al., 2021). In general, many sources of logs, e.g. from firewalls, web servers, and vulnerability scanners, might be collected and standardized before being incorporated into the learning model (Lin et al., 2009). These approaches highlight the significance of considering multiple sources of data in the learning process. However, these methods typically apply data fusion either early or late during the process of integration. While they aim to enrich the learning model with comprehensive knowledge, such fusion techniques may overlook interactions among multiple sources, potentially being less effective than *intermediate* fusion (Nayak and Luong, 2023).

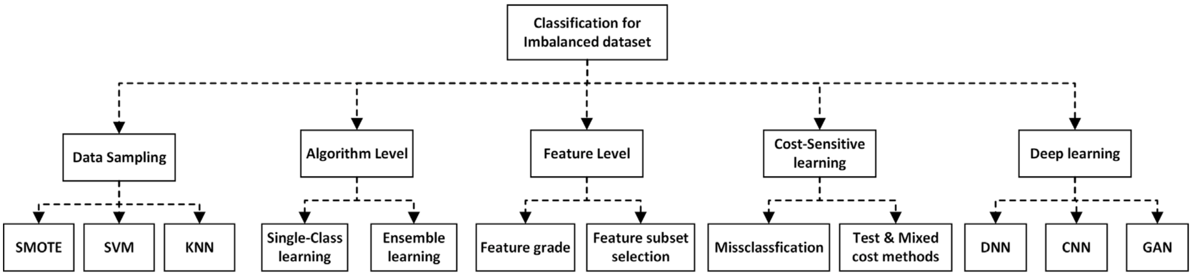

Fig. 5. The main techniques for classifying imbalanced datasets.

5.4. Challenge 4: Evolving adversarial techniques limit the effectiveness of retrospective learning

The challenge posed by adversarial techniques in cyber threat hunting lies in the ability of adversaries to adapt and evolve their tactics, developing sophisticated strategies to obfuscate their activities and circumvent traditional defense mechanisms.

Adversarial techniques employ evasion, manipulation, and vulnerability exploitation to evade detection, including zero-day exploits (Chen et al., 2022a), and advanced persistent threats (APTs) (Mahmoud et al., 2023).

Evasion Techniques: Adversaries utilize various evasion tactics to conceal their presence and activities from detection systems, including encryption, obfuscation, and stealthy manipulation of network traffic to evade IDS, IPS (intrusion prevention systems), and antivirus software. SOTA cyber threat hunting combines advanced technologies, proactive methods, and human expertise to counter evasion tactics, employing a multi-layered approach to detect and mitigate sophisticated cyber threats. ML-based threat models may be trained using adversarial examples, i.e. malicious inputs deliberately designed to deceive the model (Debicha et al., 2021; Malik et al., 2022). A threat detection model can also be strengthened against evasion techniques by learning the robust features that capture both benign and malicious data (Zhang et al., 2016). Additionally, ensemble models, which combine multiple weak classifiers to form a robust classifier, further reduce vulnerability to evasion strategies (Ahmed et al., 2022; Li and Li, 2020). However, supervised learning-based methods that rely heavily on retrospective training data remain at a disadvantage against the dynamic and ever-evolving nature of daily emerging threats.

Zero-Day Exploits: Zero-day exploits are vulnerabilities in software or systems that are unknown to vendors or security experts. Attackers exploit these vulnerabilities to launch targeted attacks without being detected by signature-based security tools. The detection and mitigation of zero-day exploits require advanced threat-hunting capabilities (Sun et al., 2018), including the use of intrusion detection, anomaly detection, and behavioral analysis.

Intrusion Detection and Prevention Systems: These systems are designed to continually monitor network traffic to detect suspicious activity. The incorporation of ML techniques such as deep learning (DL), clustering, and model updating into IDS can improve the ability to detect and respond to evolving cyber threats, including zero-day exploits. This suggests the need for multi-component or hybrid systems to improve the detection and categorization of threats. For example, a hybrid unsupervised system Pu et al. (2020) was designed to detect zero-day network intrusion attacks, combining shallow ML algorithms such as subspace clustering and support vector machine (SVM) classification to detect attacks without any prior knowledge. Similarly, a DL-based IDS framework Soltani et al. (2023) was designed for adaptive intrusion detection, utilizing multiple phases to adaptively improve threat detection and response capabilities.

Anomaly Detection and Behavior Analysis: Advanced security solutions monitor systems and networks for anomalies and unusual patterns. By analyzing the behavior of software and users, these solutions can identify potential zero-day exploits and other sophisticated attacks that traditional signature-based detection methods may overlook (Blaise et al., 2020; Dumitrasc, 2023).

5.5. Challenge 5: Scarcity of human experts and data intelligence limits threat hunting

The shortage of human experts with specialized skills and expertise is an important challenge of cybersecurity, along with the explosive development of online platforms, information technologies, and digital technologies. It poses a critical risk to organizations' ability to detect, analyze, and mitigate cyber threats, potentially leaving them vulnerable to security breaches.

The role of the cyber threat analyst emerges as paramount, particularly in the context of dynamic and increasingly sophisticated cyber threats. As adversaries adapt their TTPs, the task of identifying and validating threats becomes critical (Apruzzese et al., 2023b). Cyber threat analysts are integral to this process, employing a combination of technical expertise and analytical capabilities to sift through alerts generated by ML and DL models. Their work involves not only the identification of potential threats but also the validation of these threats to ensure accuracy and relevancy. To validate the threats, they usually rely on a broad range of sources for information on emerging threats, attacker tactics, and IoCs documents (Sauerwein et al., 2019). These sources vary widely, encompassing everything from open and commercial data feeds to services provided by threat intelligence firms (Sauerwein et al., 2019). However, as the variety and use of these threat intelligence sources expand, there remain unresolved questions about their effectiveness, especially concerning the quality of data they deliver.

To address the shortage of human experts, current threat-hunting methodologies either rely on data analytics and ML for automation or leverage external sources that offer threat intelligence capabilities. As the volume of available intelligence grows, there is an increasing demand to augment human analysts with automated tools whenever feasible. This requires not only data that is structured appropriately but also data that can be understood and processed by machines in terms of its quality (i.g., playbooks). Numerous studies have identified data quality issues as a significant hurdle in the effective exchange of threat intelligence and in information security practices due to the subpar quality of input data being utilized. Yet, despite these challenges, there is a lack of comprehensive research on data quality issues in the realm of threat hunting (Zibak et al., 2022).

Due to the complexity and dynamic nature of the threat landscape and the intelligence data quality issues, this threat hunter verification process is essential for the development of defensive playbooks where the balance between false positives and false negatives holds significant implications for the efficiency and effectiveness of cybersecurity operations team to protect organizational assets. However, the reliance on ML and DL models for threat detection presents its own set of challenges. While these technologies offer the ability to process and analyze vast quantities of data at speeds unattainable by human analysts, they also tend to generate a high volume of alerts, including both legitimate threats and benign anomalies. This creates a situation where cyber threat analysts must engage in extensive verification processes, which can be both time-consuming and labor-intensive. The need to meticulously review and corroborate each alert to distinguish between true threats and false alarms underscores the criticality of the cyber threat analyst's role. Moreover, the sheer volume of alerts, including those identifying potentially malicious groups or activities, requires not only technical acumen but also a significant investment of time and resources. This scenario underscores the importance of enhancing the efficiency of threat detection and validation processes, possibly through the integration of more sophisticated algorithms and the refinement of existing ML and DL models to reduce the incidence of false positives and negatives, thereby streamlining the workload of cyber threat analysts and improving the overall efficacy of cybersecurity measures (Gao et al., 2021a; Kaur et al., 2023).

Data Analytics and Machine Learning: This approach focuses on maximizing automation by applying data analytics and ML to minimize or even eliminate human intervention. It encompasses data collection, analysis, anomaly detection, and automated response. For instance, ML-based models can predict and trigger predefined actions like isolating or blocking viruses (Adedoyin and Teymourlouei, 2021) or malicious activities (Prabu and Sudhakar, 2023). Over the past decade, various ML models, from unsupervised (Chen et al., 2022b; Kayhan et al., 2023; Wei et al., 2020) to supervised systems (Villarreal-Vasquez et al., 2021; Fotiadou et al., 2021), have been developed to automate the threat detection process.

Threat Intelligence Integration and Orchestration: This approach involves multi-phase systems that integrate and aggregate threat intelligence from diverse sources (Wagner et al., 2019), establishing a knowledge base to support real-time threat detection and orchestration. For example, a framework for automated threat hunting was proposed (Arafune et al., 2022b) that can automatically seek out TTPs from industrial control systems based on threat intelligence provided by MITRE ATT&CK. Another automated system for threat hunting (Mavroeidis and Jøsang, 2018) analyzes Sysmon logs to categorize system processes into different threat levels based on identified characteristics. Utilizing continuously updated threat intelligence via an ontology, the system executes automated responses to indicators of compromise. Similarly, THREATRAPTOR (Gao et al., 2021b) is an automated model to extract information regarding threat behaviors (IOCs and their relationships) from unstructured open-source cyber threat intelligence (OSCTI) reports, using the extracted information to enhance threat hunting.

However, while automated cyber-threat intelligence (CTI) collection focuses on proactive threat identification and analysis, it may not capture all evolving cyber risks. Human experts bring critical thinking and adaptability to cybersecurity, enabling organizations to understand new attack techniques and effectively mitigate risks. Additionally, proactive threat hunting involves human experts actively searching for signs of malicious activities across various sources, such as the open, deep, and dark webs, using tools like VirusTotal,[1] Yara, cyber threat intelligence platforms and MISP (Malware Information Sharing Platform and Threat Sharing).[2]

Therefore, though advanced threat-hunting systems use automation to minimize human involvement, human expertise remains crucial. A balanced model integrating automated tools and human insight is essential for flexible threat detection and mitigation, enabling effective threat detection whilst also allowing human involvement in threat hunting when available.

6. Related works

Table 6 illustrates the comparison between related works in our survey papers, highlighting the contributions and differences from existing literature.

Chen et al. (2022c) provide a critical examination of current ML techniques applied to secure IoT networks against APTs. Through an extensive review, the authors identify and discuss the primary vulnerabilities of IoT systems, the nature of APTs, and the range of ML methods deployed to enhance security measures. Key ML techniques, including AdaBoost, decision trees, KNN, linear regression, random forest, support vector machines, and deep learning, are explored for their efficacy in identifying and mitigating cyber threats in the IoT landscape. Despite the potential of ML to revolutionize IoT security, the paper underscores significant challenges in effectively detecting APTs due to their stealthy characteristics and the low volume of APT traffic compared to normal network activities. This difficulty is exacerbated by the absence of comprehensive datasets encompassing the full spectrum of APT attacks, hindering the development and training of ML models tailored for APT detection in IoT contexts. The survey also emphasizes the need for future research to bridge the gap between general cyber threat detection and the nuanced requirements of APT defense, suggesting directions for advancing ML methodologies, generating APT-specific datasets, and exploring new defense mechanisms.

Qian et al. (2020) offer an exhaustive review and taxonomy of the processes involved in developing, deploying, and maintaining ML applications in the IoT domain. The paper explores various ML methods employed in IoT applications, ranging from traditional machine learning and deep learning to reinforcement learning techniques. The review scrutinizes the deployment of these models within the heterogeneous and resource-constrained environments characteristic of IoT systems, addressing the complexities of software deployment, scalability, and interoperability. It underscores the significance of addressing the unique challenges posed by the IoT environment to harness the full potential of ML technologies in this rapidly expanding field.

Rahman et al. (2023) focus on automated cyberthreat intelligence (CTI) extraction from textual sources, providing an expansive overview of current methodologies, challenges, and advancements. The paper categorizes CTI extraction purposes, identifies various textual sources of CTI, and examines the technical hurdles in automating CTI extraction. The survey emphasizes the application of NLP and ML techniques as pivotal in parsing the vast and growing corpus of textual content related to cybersecurity threats. These technologies play a crucial role in extracting actionable intelligence from unstructured data, facilitating timely decision-making processes in cybersecurity operations. While the paper does not present direct experimental results or specific datasets, it provides a critical analysis of different ML approaches employed in CTI extraction, including AdaBoost, decision trees, KNN, linear regression, random forest, SVM, and deep learning. The survey highlights the challenges posed by APTs in IoT networks, noting the difficulties in applying ML methods to detect such stealthy and low-occurrence cyber threats. The paper hypothesizes that despite the promising capabilities of ML in general cyber threat detection, significant challenges remain in accurately identifying APTs due to their sophisticated nature.

Nour et al. (2023) provide a comprehensive review of the concept of threat hunting in enterprise networks. The survey categorizes existing threat hunting solutions based on the techniques used, such as machine learning/AI-based threat hunting, graph-based threat hunting, rule-based threat hunting, and statistical-based threat hunting. The paper highlights the necessity of adopting new proactive defense approaches in response to evolving threats, where APTs utilize sophisticated techniques to remain undetected for prolonged periods. Threat hunting is presented as an iterative approach to generating and revising threat hypotheses, aiming to uncover stealthy attacks and malicious activities that standard detection mechanisms might miss. The survey underscores the role of threat hunting in improving early attack detection, leveraging various manual and automated tools/techniques to test and validate initial hypotheses, ultimately contributing to the development of more concrete and efficient threat hunting solutions for enterprise networks

[1] https://www.virustotal.com/.
[2] https://www.misp-project.org/.

Table 6
Comparison of our contribution with related surveys.

Study	Contribution	Year	Covered solutions	Threat hunting	Year Interval	Scope
Chen et al. (2022c)	• Review AI-based approaches for network intrusion detection and mitigation. • Summarize public IoT datasets for attack detection, scarcely covered in the existing literature. • Discuss opportunities and challenges in tackling APT attacks	2022	14	⊘	2007–2021	• Limited to IoT structure. • Missing the hypothesis background for threat detection. • Limited to AI techniques
Qian et al. (2020)	• Outline a pipeline for ready-to-deploy ML models and explore techniques for each stage. • Review software deployment methods and challenges specific to ML models in IoT settings.	2020	21	⊘	2010–2019	• Limited to IoT structure. • Missing the hypothesis background for threat detection. • Limited to AI techniques
Rahman et al. (2023)	• Outline the observed CTI extraction pipeline steps. • List NLP and ML techniques for CTI extraction.	2023	11	⊘	2011–2021	• Limited to CTI sources. • Missing the hypothesis background for threat detection. • Limited to AI techniques
Nour et al. (2023)	• Survey threat hunting models, methods, processes, and components. • Detail threat-hunting techniques. • Highlight challenges in threat hunting.	2023	10	●	2017–2023	• Missing the techniques to generate and analyze hypothesis background for threat detection. • Missing the method solves the challenge in Heterogeneity of the dataset and extracting data from different cyber threat sources
Sun et al. (2023)	• Summarize a six-step methodology for converting cyber info to actionable knowledge for proactive defense. • Review and analyze state-of-the-art solutions in CTI mining	2023	18	⊘	2014–2023	• Limited to CTI source. • Missing the hypothesis background for threat detection. • Missing the method solves the challenge of Heterogeneity of the dataset and extracting data from different cyber threat sources.
Our study	• Offer a taxonomy of threat-hunting techniques. • Provide an insight into techniques for proactive threat detection hypothesis generation and analysis. • Examine not only AI application challenges but also AI crime in threat hunting. • Address imbalance data structure in threat hunting. • Highlight the human element in threat hunting model.	2024	29	●	2014–2024	Achieved all four research questions RQ1–RQ4

⊘ means **Partly Yes**; ● means **Yes**.

Sun et al. (2023) provide a thorough overview of the current landscape in CTI mining. This survey demonstrates the multifaceted process of transforming diverse cybersecurity-related data sources into actionable intelligence, essential for preempting and mitigating cyber threats. Through an extensive review of state-of-the-art methodologies, the survey reveals how advanced machine learning and NLP techniques can extract valuable CTI from vast quantities of unstructured data. The discussed ML methods encompass both supervised and unsupervised learning paradigms, as well as cutting-edge deep learning architectures like graph neural networks and long short-term memory networks, tailored for the nuanced requirements of CTI extraction from textual data. Although the survey does not present new experiment results, it compiles and assesses various datasets and ML approaches used in contemporary CTI mining research, offering insights into the effectiveness of these methods in identifying cybersecurity entities, events, TTPs, hacker profiles, IoCs, vulnerabilities, and malware characteristics.

7. Discussion and findings

In this section, we reflect on the findings presented in the previous section to recommend best practices and highlight areas for future research.

7.1. Reshaping the cybersecurity landscape: The impact of AI on threat hunting and cybercrime

The integration of AI into cybersecurity is revolutionizing the way organizations detect threats and defend against cyberattacks. AI's ability to analyze vast amounts of data with speed and precision offers a game-changing advantage to cybersecurity teams. This capability is particularly valuable in an era where the volume of data is exploding and cyber threats are becoming increasingly sophisticated.

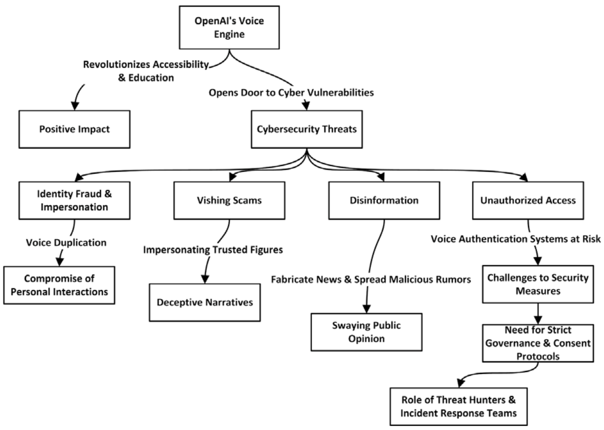
OpenAI's Voice Engine
Revolutionizes Accessibility & Education
Opens Door to Cyber Vulnerabilities
Positive Impact
Cybersecurity Threats
Identity Fraud & Impersonation
Vishing Scams
Disinformation
Unauthorized Access
Voice Duplication
Impersonating Trusted Figures
Voice Authentication Systems at Risk
Compromise of Personal Interactions
Fabricate News & Spread Malicious Rumors
Challenges to Security Measures
Deceptive Narratives
Swaying Public Opinion
Need for Strict Governance & Consent Protocols
Role of Threat Hunters & Incident Response Teams

Fig. 6. OpenAI's voice engine threat classifications.

AI is also being utilized to empower cybersecurity professionals with advanced tools for penetration testing, automating tasks that would typically require significant manual effort. Such tools can simulate attacks on a network to identify vulnerabilities before they can be exploited by malicious actors.

The potential risks arise when OpenAI's Voice Engine (OpenAI, 2024) marks a monumental stride in the field of artificial intelligence, one that mirrors human vocal characteristics with astonishing precision from merely a 15-s sample. This breakthrough is poised to revolutionize sectors such as accessibility, enhancing the user experience for those with visual impairments, and education, by offering personalized learning experiences through voice interfaces. However, the flip side of this innovation is the ushering in of considerable cybersecurity threats that necessitate vigilant examination by both entities and individuals. The potential for this technology to serve as a tool for accessibility and educational advancements illustrates its transformative power, yet it concurrently opens the door to a spectrum of cyber vulnerabilities that could exploit the very essence of personal identity (i.e., voice recognition).

Of the enumerated cyber risks, identity fraud, and impersonation stand out, highlighting a future where voice duplication could compromise the authenticity of personal interactions. Cybercriminals, armed with the capability to accurately mimic voices, could bypass voice-operated security systems or deceive individuals through social engineering tactics, thus eroding trust in voice communication. Moreover, the advent of the voice engine intensifies the threat of **vishing scams**, enabling perpetrators to craft highly convincing deceitful narratives by impersonating trusted figures. For example, Chief Information Officers call the team to shut down a server for security reasons. The technology's misuse could extend to the realm of disinformation, where voice replicas are utilized to ***fabricate news*** or ***spread malicious rumors***, potentially swaying public opinion and disrupting democratic processes. Additionally, the sanctity of voice authentication systems is at risk, as the technology could allow ***unauthorized access to sensitive information***, challenging the reliability of voice-based security measures. Lastly, the ethical and legal quandaries posed by the unauthorized use or creation of voice replicas beckon a thorough legal and moral examination, underscoring the need for strict governance and consent protocols in the deployment of such advanced AI capabilities. This will elevate the role of threat hunters and incident response teams to the next level, increasing the job's difficulty. Fig. 6 shows the threat classification of Open AI's voice engine.

7.2. *The critical role of hypothesis development in threat hunting*

Central to effective threat hunting is the development of a robust hypothesis at the outset of the research. This foundational step is critical, as it guides the systematic approach required to navigate the vast and varied nature of cyber threats.

Challenges in Current Research Approaches:

A review of the existing literature reveals a concerning trend: many research methodologies deprioritize hypothesis development in the early stages of investigation. This oversight can severely impact the efficacy of threat detection mechanisms. In the context of ransomware attacks, for example, the importance of hypothesis-driven research becomes evident. Effective threat hunting for ransomware necessitates a focus on the entire lifecycle of the attack, from the initial dumping of malicious code onto endpoints to the exploitation of vulnerabilities via command-and-control (C2) protocols.

Towards an Autonomous Threat Hunting Mechanism: The development of an autonomous threat hunting mechanism that can adapt to new and evolving tactics without prior explicit knowledge represents the frontier of cybersecurity research. Such a system must be underpinned by a robust hypothesis development process and capable of learning from a wide array of adversarial documentation. This approach is not only necessary for the advancement of threat detection models but is also critical in ensuring the resilience of cybersecurity defenses against the next generation of cyber threats.

8. Conclusion

Our comprehensive survey of 117 selected papers revealed significant insights and gaps in the academic literature on threat hunting. While traditional cybersecurity measures remain fundamental, our research highlights the critical role of threat hunting as a proactive defense mechanism against sophisticated cyber threats. This paper demonstrated that integrating AI-driven models and advanced machine learning techniques into threat hunting processes offers substantial advantages, enabling the early identification of potential threats before they escalate into major incidents. Approaches incorporating both supervised and unsupervised learning, as well as graph knowledge, have distinct advantages in the cybersecurity arsenal.

Despite the promising potential of these approaches, several challenges persist. Our review highlighted the scarcity of high-quality labeled data, the complexity of integrating multiple data sources, and the rapid evolution of adversarial techniques. Furthermore, the need for specialized expertise and standardized methodologies remains a significant hurdle. Addressing these limitations is critical for advancing cyber threat mitigation strategies, stressing the importance of continuous innovation in threat hunting methodologies to confront modern cyber threats effectively.

Our systematic review also emphasized the importance of hypothesis-driven approaches and iterative detection methodologies. We explored the distinction between threat hunting and anomaly detection, clarifying the systematic processes essential for effective threat hunting. We provide a comprehensive overview of threat hunting practices through various threat hunting datasets, analysis of supervised and unsupervised machine learning approaches, and graph-based and rule-based methods. More importantly, it included detailed tables, figures, and comparisons to highlight each's strengths and weaknesses.

In future work, we will explore deeper into the role of supervised and unsupervised models in threat hunting. We will investigate the development of standardized methodologies to ensure consistent and reliable processes across the cybersecurity community, while also focusing on improving data quality and availability. Integrating diverse data sources, such as network logs and threat intelligence feeds, will enhance our ability to detect threats. Adapting to evolving threats requires continuous innovation. This includes using emerging technologies such as quantum computing and advanced cryptography to refine hypothesis-driven approaches. These advancements are crucial for developing robust threat hunting methodologies that counter modern cyber threats.

CRediT authorship contribution statement

Arash Mahboubi: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Khanh Luong:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hamed Aboutorab:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hang Thanh Bui:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Geoff Jarrad:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Investigation, Formal analysis. **Mohammed Bahutair:** Writing – review & editing, Methodology, Investigation. **Seyit Camtepe:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Ganna Pogrebna:** Writing – review & editing, Resources, Project administration. **Ejaz Ahmed:** Writing – review & editing, Methodology. **Bazara Barry:** Writing – review & editing, Investigation. **Hannah Gately:** Writing – review & editing, Methodology.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Arash Mahboubi reports financial support was provided by Charles Sturt University. Arash Mahboubi reports a relationship with Cyber Security Cooperative Research Centre that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

The work has been supported by the Cyber Security Research Centre Limited (Grant number C22-00255) whose activities are partially funded by the Australian Government's Cooperative Research Centre Program.

References

Abdel-Basset, M., Hawash, H., Sallam, K., 2022. Federated threat-hunting approach for microservice-based industrial cyber-physical system. IEEE Trans. Ind. Inform. 18 (3), 1905–1917. http://dx.doi.org/10.1109/TII.2021.3091150.

Abououf, M., Mizouni, R., Singh, S., Otrok, H., Damiani, E., 2022. Self-supervised online and lightweight anomaly and event detection for IoT devices. IEEE Internet Things J. 9 (24), 25285–25299.

Abrams, L., 2024. New IDAT loader version uses steganography to push remcos RAT. URL: https://www.bleepingcomputer.com/news/security/new-idat-loader-version-uses-steganography-to-push-remcos-rat/. Bleeping Computer.

Abu Talib, M., Nasir, Q., Bou Nassif, A., Mokhamed, T., Ahmed, N., Mahfood, B., 2022. APT beaconing detection: A systematic review. Comput. Secur. 122, 102875. http://dx.doi.org/10.1016/j.cose.2022.102875, URL: https://www.sciencedirect.com/science/article/pii/S0167404822002693.

Adams, S., Carter, B., Fleming, C., Beling, P.A., 2018. Selecting system specific cybersecurity attack patterns using topic modeling. In: 2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering. TrustCom/BigDataSE, IEEE, pp. 490–497.

Adedoyin, A., Teymourlouei, H., 2021. Methods for automating threat hunting and response. In: 2021 International Conference on Electrical, Computer and Energy Technologies. ICECET, IEEE, pp. 1–6.

Agarwal, A., Sharma, P., Alshehri, M., Mohamed, A.A., Alfarraj, O., 2021. Classification model for accuracy and intrusion detection using machine learning approach. PeerJ Comput. Sci. 7, e437.

Aghamohammadpour, A., Mahdipour, E., Attarzadeh, I., 2023. Architecting threat hunting system based on the DODAF framework. J. Supercomput. 79 (4), 4215–4242.

Ahmed, U., Lin, J.C.-W., Srivastava, G., 2022. Mitigating adversarial evasion attacks of ransomware using ensemble learning. Comput. Electr. Eng. 100, 107903.

Akbani, R., Kwek, S., Japkowicz, N., 2004. Applying support vector machines to imbalanced datasets. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (Eds.), Machine Learning: ECML 2004. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 39–50.

Al-mamory, S.O., Algelal, Z.M., 2017. A modified DBSCAN clustering algorithm for proactive detection of DDoS attacks. In: 2017 Annual Conference on New Trends in Information & Communications Technology Applications. NTICT, IEEE, pp. 304–309.

Alevizos, L., Dekker, M., 2024. Towards an AI-enhanced cyber threat intelligence processing pipeline. arXiv preprint arXiv:2403.03265.

Almohannadi, H., Awan, I., Al Hamar, J., Cullen, A., Disso, J.P., Armitage, L., 2018. Cyber threat intelligence from honeypot data using elasticsearch. In: 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications. AINA, pp. 900–906. http://dx.doi.org/10.1109/AINA.2018.00132.

Alsaheel, A., Nan, Y., Ma, S., Yu, L., Walkup, G., Celik, Z.B., Zhang, X., Xu, D., 2021. ATLAS: A sequence-based learning approach for attack investigation. In: 30th USENIX Security Symposium. USENIX Security 21, USENIX Association, pp. 3005–3022, URL: https://www.usenix.org/conference/usenixsecurity21/presentation/alsaheel.

Alzaabi, F.R., Mehmood, A., 2024. A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods. IEEE Access 12, 30907–30927. http://dx.doi.org/10.1109/ACCESS.2024.3369906.

Anjum, M.M., Iqbal, S., Hamelin, B., 2022. ANUBIS: A provenance graph-based framework for advanced persistent threat detection. In: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing. SAC '22, Association for Computing Machinery, New York, NY, USA, pp. 1684–1693. http://dx.doi.org/10.1145/3477314.3507097.

Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A.V., Di Franco, F., 2023a. The role of machine learning in cybersecurity. Digit. Threats: Res. Pract. 4 (1), 1–38.

Apruzzese, G., Laskov, P., Montes de Oca, E., Mallouli, W., Brdalo Rapa, L., Grammatopoulos, A.V., Di Franco, F., 2023b. The role of machine learning in cybersecurity. Digit. Threats 4 (1), http://dx.doi.org/10.1145/3545574.

Arafune, M., Rajalakshmi, S., Jaldon, L., Jadidi, Z., Pal, S., Foo, E., Venkatachalam, N., 2022a. Design and development of automated threat hunting in industrial control systems. In: 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events. PerCom Workshops, pp. 618–623. http://dx.doi.org/10.1109/PerComWorkshops53856.2022.9767375.

Arafune, M., Rajalakshmi, S., Jaldon, L., Jadidi, Z., Pal, S., Foo, E., Venkatachalam, N., 2022b. Design and development of automated threat hunting in industrial control systems. In: 2022 IEEE International Conference on Pervasive Computing and Communications Workshops and Other Affiliated Events. PerCom Workshops, IEEE, pp. 618–623.

Bae, C., Tao, G., Zhang, Z., Zhang, X., 2024. Threat behavior textual search by attention graph isomorphism. In: Graham, Y., Purver, M. (Eds.), Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, St. Julian's, Malta, pp. 2616–2630, URL: https://aclanthology.org/2024.eacl-long.160.

Berady, A., Jaume, M., Tong, V.V.T., Guette, G., 2021. From TTP to IoC: Advanced persistent graphs for threat hunting. IEEE Trans. Netw. Serv. Manag. 18 (2), 1321–1333. http://dx.doi.org/10.1109/TNSM.2021.3056999.

Bhattarai, B., Huang, H., 2022. SteinerLog: Prize collecting the audit logs for threat hunting on enterprise network. In: Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security. In: ASIA CCS '22, Association for Computing Machinery, New York, NY, USA, pp. 97–108. http://dx.doi.org/10.1145/3488932.3523261.

Bibi, I., Akhunzada, A., Kumar, N., 2023. Deep AI-powered cyber threat analysis in IIoT. IEEE Internet Things J. 10 (9), 7749–7760. http://dx.doi.org/10.1109/JIOT.2022.3229722.

Blaise, A., Bouet, M., Conan, V., Secci, S., 2020. Detection of zero-day attacks: An unsupervised port-based approach. Comput. Netw. 180, 107391.

Botacin, M., Grégio, A., Alves, M.A.Z., 2021. Near-memory & in-memory detection of fileless malware. In: Proceedings of the International Symposium on Memory Systems. MEMSYS '20, Association for Computing Machinery, New York, NY, USA, pp. 23–38. http://dx.doi.org/10.1145/3422575.3422775.

Bowman, B., Laprade, C., Ji, Y., Huang, H.H., 2020. Detecting lateral movement in enterprise computer networks with unsupervised graph AI. In: 23rd International Symposium on Research in Attacks, Intrusions and Defenses. RAID 2020, USENIX Association, San Sebastian, pp. 257–268, URL: https://www.usenix.org/conference/raid2020/presentation/bowman.

Bromander, S., Jøsang, A., Eian, M., 2016. Semantic cyberthreat modelling. STIDS 74–78.

Bromander, S., Swimmer, M., Muller, L.P., Jøsang, A., Eian, M., Skjøtskift, G., Borg, F., 2021. Investigating sharing of cyber threat intelligence and proposing a new data model for enabling automation in knowledge representation and exchange. Digit. Threats 3 (1), http://dx.doi.org/10.1145/3458027.

Brown, S., Gommers, J., Serrano, O., 2015. From cyber security information sharing to threat management. In: Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security. WISCS '15, Association for Computing Machinery, New York, NY, USA, pp. 43–49. http://dx.doi.org/10.1145/2808128.2808133.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., et al., 2023. A survey on evaluation of large language models. ACM Trans. Intell. Syst. Technol..

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Chen, C.-K., Lin, S.-C., Huang, S.-C., Chu, Y.-T., Lei, C.-L., Huang, C.-Y., 2022a. Building machine learning-based threat hunting system from scratch. Digit. Threats 3 (3), http://dx.doi.org/10.1145/3491260.

Chen, C.-K., Lin, S.-C., Huang, S.-C., Chu, Y.-T., Lei, C.-L., Huang, C.-Y., 2022b. Building machine learning-based threat hunting system from scratch. Digit. Threats: Res. Pract. (DTRAP) 3 (3), 1–21.

Chen, Z., Liu, J., Shen, Y., Simsek, M., Kantarci, B., Mouftah, H.T., Djukic, P., 2022c. Machine learning-enabled IoT security: Open issues and challenges under advanced persistent threats. ACM Comput. Surv. 55 (5), http://dx.doi.org/10.1145/3530812.

Costa, B., Bachiega, J., de Carvalho, L.R., Araujo, A.P.F., 2022. Orchestration in fog computing: A comprehensive survey. ACM Comput. Surv. 55 (2), http://dx.doi.org/10.1145/3486221.

CrowdStrike, 2023. What is cyber threat hunting? https://www.crowdstrike.com/cybersecurity-101/threat-hunting/. (Accessed 14 March 2024).

CrowdStrike, 2024. 2024 Global threat report. URL: https://www.crowdstrike.com/global-threat-report/.

Debicha, I., Debatty, T., Dricot, J.-M., Mees, W., 2021. Adversarial training for deep learning-based intrusion detection systems. arXiv preprint arXiv:2104.09852.

Dekel, L., Leybovich, I., Zilberman, P., Puzis, R., 2023. MABAT: A multi-armed bandit approach for threat-hunting. IEEE Trans. Inf. Forensics Secur. 18, 477–490. http://dx.doi.org/10.1109/TIFS.2022.3215010.

Dong, C., Yang, J., Liu, S., Wang, Z., Liu, Y., Lu, Z., 2023. C-BEDIM and S-BEDIM: Lateral movement detection in enterprise network through behavior deviation measurement. Comput. Secur. 130, 103267. http://dx.doi.org/10.1016/j.cose.2023.103267, URL: https://www.sciencedirect.com/science/article/pii/S0167404823001773.

Dritsoula, L., Loiseau, P., Musacchio, J., 2017. A game-theoretic analysis of adversarial classification. IEEE Trans. Inf. Forensics Secur. 12 (12), 3094–3109. http://dx.doi.org/10.1109/TIFS.2017.2718494.

Du, M., Li, F., Zheng, G., Srikumar, V., 2017a. DeepLog: Anomaly detection and diagnosis from system logs through deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. CCS '17, Association for Computing Machinery, New York, NY, USA, pp. 1285–1298. http://dx.doi.org/10.1145/3133956.3134015.

Du, M., Li, F., Zheng, G., Srikumar, V., 2017b. Deeplog: Anomaly detection and diagnosis from system logs through deep learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. pp. 1285–1298.

Dumitrasc, V., 2023. Anomaly Detection Through User Behaviour Analysis (Master's thesis). Universitat Politécnica de Catalunya.

Dzhanashia, K., Evsutin, O., 2024. Neural networks-based data hiding in digital images: overview. Neurocomputing 127499.

Farooq, H.M., Otaibi, N.M., 2018. Optimal machine learning algorithms for cyber threat detection. In: 2018 UKSim-AMSS 20th International Conference on Computer Modelling and Simulation. UKSim, pp. 32–37. http://dx.doi.org/10.1109/UKSim.2018.00018.

Farzad, A., Gulliver, T.A., 2020. Unsupervised log message anomaly detection. ICT Express 6 (3), 229–237.

Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Cost-sensitive learning. In: Learning from Imbalanced Data Sets. Springer International Publishing, Cham, pp. 63–78. http://dx.doi.org/10.1007/978-3-319-98074-4_4.

Fotiadou, K., Velivassaki, T.-H., Voulkidis, A., Skias, D., Tsekeridou, S., Zahariadis, T., 2021. Network traffic anomaly detection via deep learning. Information 12 (5), 215.

Gao, Y., Li, X., Peng, H., Fang, B., Yu, P.S., 2022. HinCTI: A cyber threat intelligence modeling and identification system based on heterogeneous information network. IEEE Trans. Knowl. Data Eng. 34 (2), 708–722. http://dx.doi.org/10.1109/TKDE.2020.2987019.

Gao, P., Shao, F., Liu, X., Xiao, X., Qin, Z., Xu, F., Mittal, P., Kulkarni, S.R., Song, D., 2021a. Enabling efficient cyber threat hunting with cyber threat intelligence. In: 2021 IEEE 37th International Conference on Data Engineering. ICDE, pp. 193–204. http://dx.doi.org/10.1109/ICDE51399.2021.00024.

Gao, P., Shao, F., Liu, X., Xiao, X., Qin, Z., Xu, F., Mittal, P., Kulkarni, S.R., Song, D., 2021b. Enabling efficient cyber threat hunting with cyber threat intelligence. In: 2021 IEEE 37th International Conference on Data Engineering. ICDE, IEEE, pp. 193–204.

Gelenbe, E., Nakıp, M., 2022. Traffic based sequential learning during botnet attacks to identify compromised iot devices. IEEE Access 10, 126536–126549.

Habibi Lashkari, A., Kaur, G., Rahali, A., 2020. Didarknet: A contemporary approach to detect and characterize the darknet traffic using deep image learning. In: Proceedings of the 2020 10th International Conference on Communication and Network Security. pp. 1–13.

HaddadPajouh, H., Dehghantanha, A., Khayami, R., Choo, K.-K.R., 2018. A deep recurrent neural network based approach for internet of things malware threat hunting. Future Gener. Comput. Syst. 85, 88–96. http://dx.doi.org/10.1016/j.future.2018.03.007, URL: https://www.sciencedirect.com/science/article/pii/S0167739X1732486X.

Hassan, W.U., Bates, A., Marino, D., 2020. Tactical provenance analysis for endpoint detection and response systems. In: 2020 IEEE Symposium on Security and Privacy. SP, pp. 1172–1189. http://dx.doi.org/10.1109/SP40000.2020.00096.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21 (9), 1263–1284. http://dx.doi.org/10.1109/TKDE.2008.239.

Hemberg, E., Turner, M.J., Rutar, N., O'reilly, U.-M., 2024. Enhancements to threat, vulnerability, and mitigation knowledge for cyber analytics, hunting, and simulations. Digit. Threats 5 (1), http://dx.doi.org/10.1145/3615668.

Ho, G., Dhiman, M., Akhawe, D., Paxson, V., Savage, S., Voelker, G.M., Wagner, D., 2021. Hopper: Modeling and detecting lateral movement. In: 30th USENIX Security Symposium. USENIX Security 21, USENIX Association, pp. 3093–3110, URL: https://www.usenix.org/conference/usenixsecurity21/presentation/ho.

Homayoun, S., Dehghantanha, A., Ahmadzadeh, M., Hashemi, S., Khayami, R., 2020. Know abnormal, find evil: Frequent pattern mining for ransomware threat hunting and intelligence. IEEE Trans. Emerg. Top. Comput. 8 (2), 341–351. http://dx.doi.org/10.1109/TETC.2017.2756908.

Homayoun, S., Dehghantanha, A., Ahmadzadeh, M., Hashemi, S., Khayami, R., Choo, K.-K.R., Newton, D.E., 2019. DRTHIS: Deep ransomware threat hunting and intelligence system at the fog layer. Future Gener. Comput. Syst. 90, 94–104. http://dx.doi.org/10.1016/j.future.2018.07.045, URL: https://www.sciencedirect.com/science/article/pii/S0167739X17328467.

Horta Neto, A.J., Fernandes Pereira dos Santos, A., 2020. Cyber threat hunting through automated hypothesis and multi-criteria decision making. In: 2020 IEEE International Conference on Big Data. Big Data, pp. 1823–1830. http://dx.doi.org/10.1109/BigData50022.2020.9378213.

Hossain, M.N., Milajerdi, S.M., Wang, J., Eshete, B., Gjomemo, R., Sekar, R., Stoller, S.D., Venkatakrishnan, V.N., 2017. SLEUTH: real-time attack scenario reconstruction from COTS audit data. In: Proceedings of the 26th USENIX Conference on Security Symposium. SEC '17, USENIX Association, USA, pp. 487–504.

IBM, 2023. What is threat hunting? https://www.ibm.com/topics/threat-hunting. (Accessed 14 March 2024).

IBM, 2024. Incident response: The definitive guide. https://www.ibm.com/topics/incident-response. (Accessed 06 January 2024).

Jadidi, Z., Lu, Y., 2021. A threat hunting framework for industrial control systems. IEEE Access 9, 164118–164130. http://dx.doi.org/10.1109/ACCESS.2021.3133260.

Jahromi, A.N., Hashemi, S., Dehghantanha, A., Parizi, R.M., Choo, K.-K.R., 2020. An enhanced stacked LSTM method with no random initialization for malware threat hunting in safety and time-critical systems. IEEE Trans. Emerg. Top. Comput. Intell. 4 (5), 630–640. http://dx.doi.org/10.1109/TETCI.2019.2910243.

Janjua, F., Masood, A., Abbas, H., Rashid, I., 2020. Handling insider threat through supervised machine learning techniques. Procedia Comput. Sci. 177, 64–71. http://dx.doi.org/10.1016/j.procs.2020.10.012, URL: https://www.sciencedirect.com/science/article/pii/S1877050920322778. The 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) / The 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020) / Affiliated Workshops.

Johnsen, J.W., Franke, K., 2019. The impact of preprocessing in natural language for open source intelligence and criminal investigation. In: 2019 IEEE International Conference on Big Data. Big Data, pp. 4248–4254. http://dx.doi.org/10.1109/BigData47090.2019.9006006.

Jurcut, A., Niculcea, T., Ranaweera, P., Le-Khac, N.-A., 2020. Security considerations for internet of things: A survey. SN Comput. Sci. 1 (4), 193. http://dx.doi.org/10.1007/s42979-020-00201-3.

Kaiser, F.K., Dardik, U., Elitzur, A., Zilberman, P., Daniel, N., Wiens, M., Schultmann, F., Elovici, Y., Puzis, R., 2023. Attack hypotheses generation based on threat intelligence knowledge graph. IEEE Trans. Dependable Secure Comput. 20 (6), 4793–4809. http://dx.doi.org/10.1109/TDSC.2022.3233703.

Kaloudi, N., Li, J., 2020. The AI-based cyber threat landscape: A survey. ACM Comput. Surv. 53 (1), http://dx.doi.org/10.1145/3372823.

Kaur, R., Gabrijelčič, D., Klobučar, T., 2023. Artificial intelligence for cybersecurity: Literature review and future research directions. Inf. Fusion 97, 101804. http://dx.doi.org/10.1016/j.inffus.2023.101804, URL: https://www.sciencedirect.com/science/article/pii/S1566253523001136.

Kayhan, V.O., Agrawal, M., Shivendu, S., 2023. Cyber threat detection: Unsupervised hunting of anomalous commands (UHAC). Decis. Support Syst. 168, 113928. http://dx.doi.org/10.1016/j.dss.2023.113928, URL: https://www.sciencedirect.com/science/article/pii/S0167923623000039.

Khraisat, A., Gondal, I., Vamplew, P., Kamruzzaman, J., Alazab, A., 2020. Hybrid intrusion detection system based on the stacking ensemble of c5 decision tree classifier and one class support vector machine. Electronics 9 (1), 173.

King, I.J., Huang, H.H., 2023. Euler: Detecting network lateral movement via scalable temporal link prediction. ACM Trans. Priv. Secur. 26 (3), http://dx.doi.org/10.1145/3588771.

Kleinberg, S., Mishra, B., 2012. The temporal logic of causal structures. arXiv preprint arXiv:1205.2634.

Kumar, P., Gupta, G.P., Tripathi, R., Garg, S., Hassan, M.M., 2023. DLTIF: Deep learning-driven cyber threat intelligence modeling and identification framework in IoT-enabled maritime transportation systems. IEEE Trans. Intell. Transp. Syst. 24 (2), 2472–2481. http://dx.doi.org/10.1109/TITS.2021.3122368.

Lame, G., 2019. Systematic literature reviews: An introduction. In: Proceedings of the Design Society: International Conference on Engineering Design. Vol. 1, pp. 1633–1642. http://dx.doi.org/10.1017/dsi.2019.169.

Lerner, A., Romanov, A., 2022. Embedding digital information into the audio stream of a video conference for robot remote control. In: 2022 International Conference on Industrial Engineering, Applications and Manufacturing. ICIEAM, IEEE, pp. 728–733.

Li, T., Jiang, Y., Lin, C., Obaidat, M.S., Shen, Y., Ma, J., 2023a. DeepAG: Attack graph construction and threats prediction with bi-directional deep learning. IEEE Trans. Dependable Secure Comput. 20 (1), 740–757. http://dx.doi.org/10.1109/TDSC.2022.3143551.

Li, D., Li, Q., 2020. Adversarial deep ensemble: Evasion attacks and defenses for malware detection. IEEE Trans. Inf. Forensics Secur. 15, 3886–3900.

Li, T., Liu, X., Qiao, W., Zhu, X., Shen, Y., Ma, J., 2023b. T-trace: Constructing the APTs provenance graphs through multiple syslogs correlation. IEEE Trans. Dependable Secure Comput. 1–17. http://dx.doi.org/10.1109/TDSC.2023.3273918.

Li, H., Wu, J., Xu, H., Li, G., Guizani, M., 2022a. Explainable intelligence-driven defense mechanism against advanced persistent threats: A joint edge game and AI approach. IEEE Trans. Dependable Secure Comput. 19 (2), 757–775. http://dx.doi.org/10.1109/TDSC.2021.3130944.

Li, J., Zhang, R., Liu, J., Liu, G., et al., 2022b. LogKernel: A threat hunting approach based on behaviour provenance graph and graph kernel clustering. Secur. Commun. Netw. 2022.

Lin, Y.-D., Wang, Z.-Y., Lin, P.-C., Nguyen, V.-L., Hwang, R.-H., Lai, Y.-C., 2022. Multi-datasource machine learning in intrusion detection: Packet flows, system logs and host statistics. J. Inf. Secur. Appl. 68, 103248.

Lin, Q., Zhang, H., Lou, J.-G., Zhang, Y., Chen, X., 2016. Log clustering based problem identification for online service systems. In: Proceedings of the 38th International Conference on Software Engineering Companion. pp. 102–111.

Lin, C., Zhitang, L., Cuixia, G., 2009. Automated analysis of multi-source logs for network forensics. In: 2009 First International Workshop on Education Technology and Computer Science. Vol. 1, IEEE, pp. 660–664.

Liu, Q., Stokes, J.W., Mead, R., Burrell, T., Hellen, I., Lambert, J., Marochko, A., Cui, W., 2018. Latte: Large-scale lateral movement detection. In: MILCOM 2018 - 2018 IEEE Military Communications Conference. MILCOM, pp. 1–6. http://dx.doi.org/10.1109/MILCOM.2018.8599748.

Liu, C., Wang, J., Chen, X., 2022. Threat intelligence ATT&CK extraction based on the attention transformer hierarchical recurrent neural network. Appl. Soft Comput. 122, 108826. http://dx.doi.org/10.1016/j.asoc.2022.108826, URL: https://www.sciencedirect.com/science/article/pii/S1568494622002289.

Lou, P., Lu, G., Jiang, X., Xiao, Z., Hu, J., Yan, J., 2021. Cyber intrusion detection through association rule mining on multi-source logs. Appl. Intell. 51, 4043–4057.

Mahboubi, A., Ansari, K., Camtepe, S., 2021. Using process mining to identify file system metrics impacted by ransomware execution. In: Bouzefrane, S., Laurent, M., Boumerdassi, S., Renault, E. (Eds.), Mobile, Secure, and Programmable Networking. Springer International Publishing, Cham, pp. 57–71.

Mahmoud, M., Mannan, M., Youssef, A., 2023. APTHunter: Detecting advanced persistent threats in early stages. Digit. Threats 4 (1), http://dx.doi.org/10.1145/3559768.

Malik, A.-E., Andresini, G., Appice, A., Malerba, D., 2022. An XAI-based adversarial training approach for cyber-threat detection. In: 2022 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress. DASC/PiCom/CBDCom/CyberSciTech, IEEE, pp. 1–8.

Mansfield-Devine, S., 2017. Threat hunting: assuming the worst to strengthen resilience. Netw. Secur. 2017 (5), 13–17. http://dx.doi.org/10.1016/S1353-4858(17)30050-8, URL: https://www.sciencedirect.com/science/article/pii/S1353485817300508.

Marin, E., Almukaynizi, M., Shakarian, P., 2020. Inductive and deductive reasoning to assist in cyber-attack prediction. In: 2020 10th Annual Computing and Communication Workshop and Conference. CCWC, pp. 0262–0268. http://dx.doi.org/10.1109/CCWC47524.2020.9031154.

Martins, C., Medeiros, I., 2022. Generating quality threat intelligence leveraging OSINT and a cyber threat unified taxonomy. ACM Trans. Priv. Secur. 25 (3), http://dx.doi.org/10.1145/3530977.

Mavroeidis, V., Bromander, S., 2017. Cyber threat intelligence model: An evaluation of taxonomies, sharing standards, and ontologies within cyber threat intelligence. In: 2017 European Intelligence and Security Informatics Conference. EISIC, pp. 91–98. http://dx.doi.org/10.1109/EISIC.2017.20.

Mavroeidis, V., Jøsang, A., 2018. Data-driven threat hunting using sysmon. In: Proceedings of the 2nd International Conference on Cryptography, Security and Privacy. In: ICCSP 2018, Association for Computing Machinery, New York, NY, USA, pp. 82–88. http://dx.doi.org/10.1145/3199478.3199490.

Mavroeidis, V., Jøsang, A., 2018. Data-driven threat hunting using sysmon. In: Proceedings of the 2nd International Conference on Cryptography, Security and Privacy. pp. 82–88.

Melman, A., Evsutin, O., 2024. Image watermarking based on a ratio of DCT coefficient sums using a gradient-based optimizer. Comput. Electr. Eng. 117, 109271.

Meng, W., Liu, Y., Zhu, Y., Zhang, S., Pei, D., Liu, Y., Chen, Y., Zhang, R., Tao, S., Sun, P., et al., 2019. Loganomaly: Unsupervised detection of sequential and quantitative anomalies in unstructured logs. In: IJCAI. Vol. 19, pp. 4739–4745.

Mikhail, J.W., Williams, J.C., Roelke, G.R., 2020. procmonML: Generating evasion resilient host-based behavioral analytics from tree ensembles. Comput. Secur. 98, 102002. http://dx.doi.org/10.1016/j.cose.2020.102002, URL: https://www.sciencedirect.com/science/article/pii/S0167404820302753.

Milajerdi, S.M., Eshete, B., Gjomemo, R., Venkatakrishnan, V., 2019a. POIROT: Aligning attack behavior with kernel audit records for cyber threat hunting. In: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. CCS '19, Association for Computing Machinery, New York, NY, USA, pp. 1795–1812. http://dx.doi.org/10.1145/3319535.3363217.

Milajerdi, S.M., Gjomemo, R., Eshete, B., Sekar, R., Venkatakrishnan, V., 2019b. Holmes: real-time apt detection through correlation of suspicious information flows. In: 2019 IEEE Symposium on Security and Privacy. SP, IEEE, pp. 1137–1152.

Mimoso, M., 2015. Stegoloader malware uses steganography to hide itself. URL: https://threatpost.com/information-stealing-stegoloader-malware-hides-in-images/113337/. Threatpost.

Mohamad, A.M., Yan, F.Y.Y., Aziz, N.A., Norhisham, S., 2022. Inductive-deductive reasoning in qualitative analysis using atlas.ti: Trending cybersecurity Twitter data analytics. In: 2022 3rd International Conference for Emerging Technology. INCET, pp. 1–5. http://dx.doi.org/10.1109/INCET54531.2022.9824944.

MontazeriShatoori, M., Davidson, L., Kaur, G., Lashkari, A.H., 2020. Detection of doh tunnels using time-series classification of encrypted traffic. In: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress. DASC/PiCom/CBDCom/CyberSciTech, IEEE, pp. 63–70.

Moti, Z., Hashemi, S., Jahromi, A.N., 2020. A deep learning-based malware hunting technique to handle imbalanced data. In: 2020 17th International ISC Conference on Information Security and Cryptology. ISCISC, IEEE, pp. 48–53.

Mukkamala, S., Janoski, G., Sung, A., 2002. Intrusion detection: support vector machines and neural networks. In: Proceedings of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO. pp. 1702–1707.

Nakip, M., Gelenbe, E., 2021. MIRAI botnet attack detection with auto-associative dense random neural network. In: 2021 IEEE Global Communications Conference. GLOBECOM, IEEE, pp. 01–06.

Nakıp, M., Gelenbe, E., 2024. Online self-supervised deep learning for intrusion detection systems. IEEE Trans. Inf. Forensics Secur..

Nakip, M., Gül, B.C., Gelenbe, E., 2023. Decentralized online federated g-network learning for lightweight intrusion detection. In: 2023 31st International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. MASCOTS, IEEE, pp. 1–8.

Narayanan, S.N., Ganesan, A., Joshi, K., Oates, T., Joshi, A., Finin, T., 2018. Early detection of cybersecurity threats using collaborative cognition. In: 2018 IEEE 4th International Conference on Collaboration and Internet Computing. CIC, IEEE, pp. 354–363.

Nayak, R., Luong, K., 2023. Multi-aspect Learning: Methods and Applications, vol. 242, Springer Nature.

Niakanlahiji, A., Wei, J., Alam, M.R., Wang, Q., Chu, B.-T., 2020. ShadowMove: A stealthy lateral movement strategy. In: 29th USENIX Security Symposium. USENIX Security 20, USENIX Association, pp. 559–576, URL: https://www.usenix.org/conference/usenixsecurity20/presentation/niakanlahiji.

Nour, B., Pourzandi, M., Debbabi, M., 2023. A survey on threat hunting in enterprise networks. IEEE Commun. Surv. Tutor. 25 (4), 2299–2324. http://dx.doi.org/10.1109/COMST.2023.3299519.

Oliner, A., Stearley, J., 2007. What supercomputers say: A study of five system logs. In: 37th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. DSN'07, pp. 575–584. http://dx.doi.org/10.1109/DSN.2007.103.

OpenAI, 2024. Navigating the challenges and opportunities of synthetic voices. URL: https://openai.com/blog/navigating-the-challenges-and-opportunities-of-synthetic-voices. (Accessed 8 April 2024).

Pal, P., Chattopadhyay, P., Swarnkar, M., 2023. Temporal feature aggregation with attention for insider threat detection from activity logs. Expert Syst. Appl. 224, 119925. http://dx.doi.org/10.1016/j.eswa.2023.119925, URL: https://www.sciencedirect.com/science/article/pii/S0957417423004268.

Prabu, K., Sudhakar, P., 2023. An automated intrusion detection and prevention model for enhanced network security and threat assessment. Int. J. Comput. Netw. Appl. 10 (4).

Pu, G., Wang, L., Shen, J., Dong, F., 2020. A hybrid unsupervised clustering-based anomaly detection method. Tsinghua Sci. Technol. 26 (2), 146–153.

Qian, B., Su, J., Wen, Z., Jha, D.N., Li, Y., Guan, Y., Puthal, D., James, P., Yang, R., Zomaya, A.Y., Rana, O., Wang, L., Koutny, M., Ranjan, R., 2020. Orchestrating the development lifecycle of machine learning-based IoT applications: A taxonomy and survey. ACM Comput. Surv. 53 (4), http://dx.doi.org/10.1145/3398020.

Rahman, M.R., Hezaveh, R.M., Williams, L., 2023. What are the attackers doing now? Automating cyberthreat intelligence extraction from text on pace with the changing threat landscape: A survey. ACM Comput. Surv. 55 (12), http://dx.doi.org/10.1145/3571726, URL: https://doi-org.ezproxy.csu.edu.au/10.1145/3571726.

Rashid, A.N.M.B., Ahmed, M., Sikos, L.F., Haskell-Dowland, P., 2022. Anomaly detection in cybersecurity datasets via cooperative co-evolution-based feature selection. ACM Trans. Manage. Inf. Syst. 13 (3), http://dx.doi.org/10.1145/3495165.

Rizvi, A.S.M., Bertholdo, L., Ceron, J., Heidemann, J., 2022. Anycast agility: Network playbooks to fight DDoS. In: 31st USENIX Security Symposium. USENIX Security 22, USENIX Association, Boston, MA, pp. 4201–4218, URL: https://www.usenix.org/conference/usenixsecurity22/presentation/rizvi.

Rodriguez, R., Rodriguez, J.L., 2022. The ThreatHunter-playbook. GitHub repository. https://github.com/OTRF/ThreatHunter-Playbook.

Rosli, N.A., Yassin, W., Faizal, M., Selamat, S.R., 2019. Clustering analysis for malware behavior detection using registry data. Int. J. Adv. Comput. Sci. Appl. (IJACSA) 10, 12.

Saeed, S., Suayyid, S.A., Al-Ghamdi, M.S., Al-Muhaisen, H., Almuhaideb, A.M., 2023. A systematic literature review on cyber threat intelligence for organizational cybersecurity resilience. Sensors 23 (16), http://dx.doi.org/10.3390/s23167273, URL: https://www.mdpi.com/1424-8220/23/16/7273.

Salem, A., Banescu, S., Pretschner, A., 2021. Maat: Automatically analyzing VirusTotal for accurate labeling and effective malware detection. ACM Trans. Priv. Secur. 24 (4), http://dx.doi.org/10.1145/3465361.

Samtani, S., Kantarcioglu, M., Chen, H., 2020. Trailblazing the artificial intelligence for cybersecurity discipline: A multi-disciplinary research roadmap. ACM Trans. Manage. Inf. Syst. 11 (4), http://dx.doi.org/10.1145/3430360.

Satpathi, S., Deb, S., Srikant, R., Yan, H., 2019. Learning latent events from network message logs. IEEE/ACM Trans. Netw. 27 (4), 1728–1741.

Satvat, K., Gjomemo, R., Venkatakrishnan, V., 2021. Extractor: Extracting attack behavior from threat reports. In: 2021 IEEE European Symposium on Security and Privacy. EuroS&P, pp. 598–615. http://dx.doi.org/10.1109/EuroSP51992.2021.00046.

Sauerwein, C., Pekaric, I., Felderer, M., Breu, R., 2019. An analysis and classification of public information security data sources used in research and practice. Comput. Secur. 82, 140–155. http://dx.doi.org/10.1016/j.cose.2018.12.011, URL: https://www.sciencedirect.com/science/article/pii/S0167404818304978.

Schlette, D., Caselli, M., Pernul, G., 2021a. A comparative study on cyber threat intelligence: The security incident response perspective. IEEE Commun. Surv. Tutor. 23 (4), 2525–2556. http://dx.doi.org/10.1109/COMST.2021.3117338.

Schlette, D., Empl, P., Caselli, M., Schreck, T., Pernul, G., 2023. Do you play it by the books? A study on incident response playbooks and influencing factors. In: 2024 IEEE Symposium on Security and Privacy. SP, IEEE Computer Society, 60–60.

Schlette, D., Vielberth, M., Pernul, G., 2021b. CTI-SOC2M2 – the quest for mature, intelligence-driven security operations and incident response capabilities. Comput. Secur. 111, 102482. http://dx.doi.org/10.1016/j.cose.2021.102482, URL: https://www.sciencedirect.com/science/article/pii/S0167404821003060.

Shakarian, P., Parker, A., Simari, G., Subrahmanian, V.V., 2011. Annotated probabilistic temporal logic. ACM Trans. Comput. Logic (TOCL) 12 (2), 1–44.

Shang, W., Jiang, Z.M., Hemmati, H., Adams, B., Hassan, A.E., Martin, P., 2013. Assisting developers of big data analytics applications when deploying on hadoop clouds. In: 2013 35th International Conference on Software Engineering. ICSE, pp. 402–411. http://dx.doi.org/10.1109/ICSE.2013.6606586.

Shao, Z., Yuan, S., Wang, Y., 2021. Adaptive online learning for IoT botnet detection. Inform. Sci. 574, 84–95.

Sharafaldin, I., Habibi Lashkari, A., Ghorbani, A.A., 2019. A detailed analysis of the CICIDS2017 data set. In: Mori, P., Furnell, S., Camp, O. (Eds.), Information Systems Security and Privacy. Springer International Publishing, Cham, pp. 172–188.

Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A., et al., 2018. Toward generating a new intrusion detection dataset and intrusion traffic characterization. ICISSp 1, 108–116.

Sharma, P., Parvat, T.J., 2013. Network log clustering using k-means algorithm. In: Proceedings of the Third International Conference on Trends in Information, Telecommunication and Computing. Springer, pp. 115–124.

Shashanka, M., Shen, M.-Y., Wang, J., 2016. User and entity behavior analytics for enterprise security. In: 2016 IEEE International Conference on Big Data. Big Data, pp. 1867–1874. http://dx.doi.org/10.1109/BigData.2016.7840805.

Shen, Y., Stringhini, G., 2019. ATTACK2VEC: Leveraging temporal word embeddings to understand the evolution of cyberattacks. In: 28th USENIX Security Symposium. USENIX Security 19, USENIX Association, Santa Clara, CA, pp. 905–921, URL: https://www.usenix.org/conference/usenixsecurity19/presentation/shen.

Shin, H., Shim, W., Kim, S., Lee, S., Kang, Y.G., Hwang, Y.H., 2021. #Twiti: Social listening for threat intelligence. In: Proceedings of the Web Conference 2021. WWW '21, Association for Computing Machinery, New York, NY, USA, pp. 92–104. http://dx.doi.org/10.1145/3442381.3449797.

Shu, X., Araujo, F., Schales, D.L., Stoecklin, M.P., Jang, J., Huang, H., Rao, J.R., 2018. Threat intelligence computing. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, Association for Computing Machinery, New York, NY, USA, pp. 1883–1898. http://dx.doi.org/10.1145/3243734.3243829.

Soltani, M., Ousat, B., Siavoshani, M.J., Jahangir, A.H., 2023. An adaptable deep learning-based intrusion detection system to zero-day attacks. J. Inf. Secur. Appl. 76, 103516.

Sommer, R., Paxson, V., 2010. Outside the closed world: On using machine learning for network intrusion detection. In: 2010 IEEE Symposium on Security and Privacy. IEEE, pp. 305–316.

Straub, J., 2020. Modeling attack, defense and threat trees and the cyber kill chain, ATT&CK and STRIDE frameworks as blackboard architecture networks. In: 2020 IEEE International Conference on Smart Cloud. SmartCloud, pp. 148–153. http://dx.doi.org/10.1109/SmartCloud49737.2020.00035.

Sun, X., Dai, J., Liu, P., Singhal, A., Yen, J., 2018. Using Bayesian networks for probabilistic identification of zero-day attack paths. IEEE Trans. Inf. Forensics Secur. 13 (10), 2506–2521. http://dx.doi.org/10.1109/TIFS.2018.2821095.

Sun, N., Ding, M., Jiang, J., Xu, W., Mo, X., Tai, Y., Zhang, J., 2023. Cyber threat intelligence mining for proactive cybersecurity defense: A survey and new perspectives. IEEE Commun. Surv. Tutor. 25 (3), 1748–1774. http://dx.doi.org/10.1109/COMST.2023.3273282.

Syed, Z., Padia, A., Finin, T., Mathews, L., Joshi, A., 2016. UCO: A unified cybersecurity ontology. In: Workshops at the Thirtieth AAAI Conference on Artificial Intelligence.

Tabiban, A., Zhao, H., Jarraya, Y., Pourzandi, M., Zhang, M., Wang, L., 2022. ProvTalk: Towards interpretable multi-level provenance analysis in networking functions virtualization (NFV). In: NDSS.

Tang, B., Wang, J., Yu, Z., Chen, B., Ge, W., Yu, J., Lu, T., 2022. Advanced persistent threat intelligent profiling technique: A survey. Comput. Electr. Eng. 103, 108261. http://dx.doi.org/10.1016/j.compeleceng.2022.108261, URL: https://www.sciencedirect.com/science/article/pii/S0045790622004931.

Tax, D.M., Duin, R.P., 2004. Support vector data description. Mach. Learn. 54, 45–66.

Technologies, P., 2024. SteganoAmor campaign: TA558 mass-attacking companies and public institutions all around the world. URL: https://shorturl.at/6BRiE. Positive Technologies.

Villarreal-Vasquez, M., Modelo-Howard, G., Dube, S., Bhargava, B., 2021. Hunting for insider threats using LSTM-based anomaly detection. IEEE Trans. Dependable Secure Comput. 20 (1), 451–462.

Vinayakumar, R., Soman, K., Poornachandran, P., 2017. Applying convolutional neural network for network intrusion detection. In: 2017 International Conference on Advances in Computing, Communications and Informatics. ICACCI, IEEE, pp. 1222–1228.

Wagner, C., Dulaunoy, A., Wagener, G., Iklody, A., 2016. 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. WISCS '16, Association for Computing Machinery, New York, NY, USA, pp. 49–56. http://dx.doi.org/10.1145/2994539.2994542.

Wagner, T.D., Mahbub, K., Palomar, E., Abdallah, A.E., 2019. Cyber threat intelligence sharing: Survey and research directions. Comput. Secur. 87 (C), http://dx.doi.org/10.1016/j.cose.2019.101589.

Wang, Y., Bashar, M.A., Chandramohan, M., Nayak, R., 2023. Exploring topic models to discern cyber threats on Twitter: A case study on Log4Shell. Intell. Syst. Appl. 20, 200280. http://dx.doi.org/10.1016/j.iswa.2023.200280, URL: https://www.sciencedirect.com/science/article/pii/S2667305323001059.

Wang, H., Cao, Z., Hong, B., 2020. A network intrusion detection system based on convolutional neural network. J. Intell. Fuzzy Systems 38 (6), 7623–7637.

Wang, L., Han, M., Li, X., Zhang, N., Cheng, H., 2021. Review of classification methods on unbalanced data sets. IEEE Access 9, 64606–64628. http://dx.doi.org/10.1109/ACCESS.2021.3074243.

Wang, S., Wang, Z., Zhou, T., Sun, H., Yin, X., Han, D., Zhang, H., Shi, X., Yang, J., 2022a. threaTrace: Detecting and tracing host-based threats in node level through provenance graph learning. IEEE Trans. Inf. Forensics Secur. 17, 3972–3987. http://dx.doi.org/10.1109/TIFS.2022.3208815.

Wang, J., Zhao, C., He, S., Gu, Y., Alfarraj, O., Abugabah, A., 2022b. Loguad: log unsupervised anomaly detection based on word2vec. Comput. Syst. Sci. Eng. 41 (3), 1207.

Wei, R., Cai, L., Zhao, L., Yu, A., Meng, D., 2021. DeepHunter: A graph neural network based approach for robust cyber threat hunting. In: Garcia-Alfaro, J., Li, S., Poovendran, R., Debar, H., Yung, M. (Eds.), Security and Privacy in Communication Networks. Springer International Publishing, Cham, pp. 3–24.

Wei, Y., Chow, K.-P., Yiu, S.-M., 2020. Insider threat detection using multi-autoencoder filtering and unsupervised learning. In: Advances in Digital Forensics XVI: 16th IFIP WG 11.9 International Conference, New Delhi, India, January 6–8, 2020, Revised Selected Papers 16. Springer, pp. 273–290.

Woźniak, M., Siłka, J., Wieczorek, M., Alrashoud, M., 2020. Recurrent neural network model for IoT and networking malware threat detection. IEEE Trans. Ind. Inform. 17 (8), 5583–5594.

Yazdinejad, A., Dehghantanha, A., Parizi, R.M., Hammoudeh, M., Karimipour, H., Srivastava, G., 2022. Block hunter: Federated learning for cyber threat hunting in blockchain-based IIoT networks. IEEE Trans. Ind. Inform. 18 (11), 8356–8366. http://dx.doi.org/10.1109/TII.2022.3168011.

Yazdinejad, A., Kazemi, M., Parizi, R.M., Dehghantanha, A., Karimipour, H., 2023. An ensemble deep learning model for cyber threat hunting in industrial internet of things. Digit. Commun. Netw. 9 (1), 101–110. http://dx.doi.org/10.1016/j.dcan.2022.09.008, URL: https://www.sciencedirect.com/science/article/pii/S2352864822001833.

Yin, C., Zhu, Y., Fei, J., He, X., 2017. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access 5, 21954–21961.

Yousef, W.A., Traoré, I., Briguglio, W., 2021. UN-AVOIDS: Unsupervised and nonparametric approach for visualizing outliers and invariant detection scoring. IEEE Trans. Inf. Forensics Secur. 16, 5195–5210. http://dx.doi.org/10.1109/TIFS.2021.3125608.

Zang, X., Gong, J., Wang, M., Gao, P., Zhang, G., 2023a. IP traffic behavior characterization via semantic mining. J. Netw. Comput. Appl. 213, 103603. http://dx.doi.org/10.1016/j.jnca.2023.103603, URL: https://www.sciencedirect.com/science/article/pii/S108480452300022X.

Zang, X., Gong, J., Zhang, X., Li, G., 2023b. Attack scenario reconstruction via fusing heterogeneous threat intelligence. Comput. Secur. 133, 103420. http://dx.doi.org/10.1016/j.cose.2023.103420, URL: https://www.sciencedirect.com/science/article/pii/S0167404823003309.

Zhang, H., Cai, L., Zhao, L., Yu, A., Ma, J., Meng, D., 2022. LogMiner: A system audit log reduction strategy based on behavior pattern mining. In: MILCOM 2022 - 2022 IEEE Military Communications Conference. MILCOM, pp. 292–297. http://dx.doi.org/10.1109/MILCOM55135.2022.10017626.

Zhang, F., Chan, P.P.K., Biggio, B., Yeung, D.S., Roli, F., 2016. Adversarial feature selection against evasion attacks. IEEE Trans. Cybern. 46 (3), 766–777. http://dx.doi.org/10.1109/TCYB.2015.2415032.

Zibak, A., Sauerwein, C., Simpson, A.C., 2022. Threat intelligence quality dimensions for research and practice. Digit. Threats 3 (4), http://dx.doi.org/10.1145/3484202.

Dr Arash Mahboubi is a deputy leader of the Cyber Security Research Group at Charles Sturt University. He is an esteemed expert in the realm of information security, having earned his Ph.D. from Queensland University of Technology. With a deep-rooted passion for cybersecurity, Dr. Mahboubi's research extensively delves into areas such as the interplay of artificial intelligence and machine learning with cybersecurity, secure automation in cyber environments, and adaptive defense strategies against intricate cyber-attacks.

Dr Khanh Luongis a research fellow in cybersecurity at Charles Sturt University, Australia. She received her Ph.D. degree in Computer Science from Queensland University of Technology (QUT) in Australia in 2019. Her current research interests include cyber threat detection, text mining, social media data mining and multi-aspect data representation learning.

Dr Hamed Aboutorab is a cybersecurity research fellow at Charles Sturt University. He completed his Ph.D. at the University of New South Wales. His research is centered around the Integration of cybersecurity and artificial intelligence, with a current focus on security automation and orchestration. His research has been published in prestigious international journals, including IEEE Transactions on Services Computing, Expert Systems with Applications, Journal of Network and Computer Applications, and Future Generation Computer Systems.

Dr Hang Thanh Bui is a researcher at Charles Sturt University, School of Computing, Mathematics and Engineering. She has 5 years of experience working on industrial projects in the application of digitalization and cutting-edge technology. Her research interest is in cyber security, and the application of machine learning in interdisciplinary, distributed computing.

Dr Geoff Jarrad is a senior data scientist and research engineer in CSIRO Data61's Cybersecurity Automation and Orchestration team, involved with research and software engineering for AI Security. He has a long-term background in statistical machine learning, reasoning under uncertainty, and applied decision support systems. Currently, Geoff is working on Threat Hunting projects using AI/ML to automatically detect attacks on computer systems. Previous projects include: the detection of role/access conflicts for identity and access management (IAM) within the Australian banking system; and development on the StellarGraph ML library for data science approaches to anti-money laundering, spammer detection and other graph-analytic problems.

Dr Mohammed Bahutair is a postdoctoral fellow at Data61. He holds a bachelor, masters degree, and Ph.D. in Computer Engineering. His research interests include cyber security, machine learning, Internet of Things, and trust.

Dr Seyit Camtepe is a principal research scientist at CSIRO's Data61 leading the Autonomous Security and Software Security team. He is passionate about discovering unusual solutions to challenging cybersecurity problems with a specific focus on pervasive security. He was among the first to inform society about Android malware outbreak, and to realize the model-to-data paradigm in computing to enable research on data in captivity. From 2007 to 2013, he was with the TU-Berlin, Germany, as a senior researcher and research group leader in security. Dr Seyit worked for five years as an ECARD lecturer at the QUT, Australia.

Professor Ganna Pogrebna is the Executive Director of Artificial Intelligence and Cyber Futures Institute at Charles Sturt University. Blending behavioral science, AI, computer science, data analytics, engineering, and business model innovation, Ganna helps cities, businesses, charities, and individuals to better understand why they make decisions they make and how they can optimize their behavior to achieve higher profit, better social outcomes, as well as flourish and bolster their wellbeing. Ganna's recent projects focus on smart technological and social systems, cybersecurity, human–computer and human-data interactions and business models.

Dr Ejaz Ahmed holds a B.Sc. in computer sciences from Peshawar University, an M.S. from NUST Pakistan, and a Ph.D. from Kyung Hee University, South Korea. Currently a research scientist in cybersecurity at CSIRO's Data61, he previously held positions at POSTECH and Sungkyunkwan University in South Korea. With expertise in malware detection, threat hunting, digital forensics, program analysis, natural language processing, and machine learning, Ahmed is actively involved in advancing cybersecurity research.

Dr Bazara Barry, a globally recognized cyber security expert with 10+ years of leadership in ICT, emphasizes a holistic business-centric approach to cyber security. His expertise spans governance, risk assurance, compliance, project management, and R&D. As a distinguished speaker, he has published 30 peer-reviewed articles, presented on 5 continents, and won 4 international best paper awards. Notably, he received the International Telecommunication Union ITU award for proposing a Secure E-service Platform, showcasing his impactful contributions to the field.

Hannah Gately has received her Master's degree from Macquarie University, Australia with a focus on Security Studies and Criminology.