Expert Insights into Advanced Persistent Threats: Analysis, Attribution, and Challenges

Aakanksha Saha
Technische Universität Wien
aakanksha@seclab.wien

James Mattei
Tufts University
james.mattei@tufts.edu

Jorge Blasco
Univ. Politécnica de Madrid
jorge.blasco.alis@upm.es

Lorenzo Cavallaro
University College London
l.cavallaro@ucl.ac.uk

Daniel Votipka
Tufts University
dvotipka@cs.tufts.edu

Martina Lindorfer
Technische Universität Wien
martina@seclab.wien

Abstract

Advanced Persistent Threats (APTs) are sophisticated and targeted threats that demand significant effort from analysts for detection and attribution. Researchers have developed various techniques to support these efforts. However, security practitioners' perceptions and challenges in analyzing APT-level threats are not yet well understood. To address this gap, we conducted semi-structured interviews with 15 security practitioners across diverse roles and expertise. From the interview responses, we identify a three-layer approach to APT attribution, each having its own goals and challenges. We find that practitioners typically prioritize understanding the adversary's tactics, techniques, procedures (TTPs), and motivations over identifying the specific entity behind an attack. We also find challenges in existing tools and processes mostly stemming from their inability to handle diverse and complex data and issues with both internal and external collaboration. Based on these findings, we provide four recommendations for improving attribution approaches and discuss how these improvements can address the identified challenges.

1 Introduction

Advanced Persistent Threats (APTs) have become a critical instrument of modern geopolitical warfare, allowing nation-states to conduct sophisticated cyber espionage and strategic intelligence. Cyber threat analysts regularly uncover APT campaigns targeting government agencies and private sector companies [15, 17, 65]. Attribution of these campaigns has exposed evolving and sophisticated adversaries that engage in espionage, theft of information, and disruption of services. In response, researchers and industry practitioners have advanced APT detection [20, 25, 26, 30, 36, 40] and attribution [56, 58, 59, 60, 69] emphasizing its critical role in informing defensive strategies and understanding adversarial behaviors. Despite these advancements, the majority of existing research is concentrated on developing technical solutions for robust and accurate attribution. This includes automating malware clustering and leveraging machine learning for detection and attribution. However, technical solutions often do not fully engage with the professionals who actively use, manage, or attribute malware in real-world scenarios.

Understanding real-world practices is crucial for bridging the gap between theoretical models and practical applications. By examining how APT investigations are conducted in the field, we can ensure that tools and techniques are designed to meet the actual needs of analysts, align with their processes, and identify key assumptions that might simplify tool development. While previous studies have focused on reverse engineering [67] and malware analysis [74], and recent research has explored broader threat-hunting practices [8, 44], our research uniquely identifies the nuanced challenges specific to APT incidents and attribution, examining how practitioners navigate these complex scenarios.

In the area of generic, i.e., non-targeted, malware, Wong et al. [72] recently identified a significant misalignment between the practical challenges faced by malware experts and the focus of existing research solutions. Building on their observations, we explore the complexities of APTs and their attribution to better understand the disconnect between research and practical applications. Our study provides insights into the relevance and effectiveness of APT attack attribution tools and methodologies, aiming to offer a deeper understanding of *'why attribution is important'* and *'how attribution is performed'* in real-world scenarios. With this in mind, we seek to answer the following three main research questions:

RQ1 Why is attribution important, and what objectives does it serve in the context of security incidents?

RQ2 What are the key steps and processes involved in investigating and attributing APT incidents?

RQ3 What challenges and obstacles do practitioners face when investigating and attributing APT activities?

To address these questions, we conducted semi-structured interviews with a diverse group of 15 security practitioners including malware analysts, threat intelligence researchers, security consultants, and incident responders. Each of these roles plays a crucial part in the attribution process at different levels, highlighting the collaborative nature of effective attack investigation and attribution. Incident responders often initiate investigations and provide critical insights into active threats, while malware analysts and threat intelligence researchers offer detailed analyses of attack artifacts. Security Operation Center (SOC) and incident response team leads coordinate investigations, while upper management integrates findings into broader organizational strategies.

In our interviews, we explored how security practitioners investigate and attribute APTs, focusing on the tools and techniques they use for analyzing malicious samples, threat hunting, and addressing challenges in the attribution process. We also explored the use of internal and external intelligence for tracking and correlating threats, as well as broader organizational and policy-related aspects, such as collaboration between teams or agencies. The interviews provided a comprehensive overview of the strategies employed by security professionals across various roles.

We found that attribution can generally be modeled across three distinct decision levels. (1) *APT Classification*, (2) *Tactics, Techniques, and Procedures (TTP) Attribution*, and (3) *Country Attribution*. Victim organizations progress through these levels depending on the nature of the incident and their specific needs. APT classification helps differentiate between generic threats and advanced threats, thus prioritizing resources for mitigation efforts. TTP attribution involves a comprehensive and collaborative investigation to identify the specific TTPs used by the threat actor, aiding precise and effective response strategies. In contrast, country attribution—aiming to identify the exact entity (nation or country) behind the attack—is often challenging and less emphasized. Participants prioritize identifying *what* threat actors are likely to do rather than focusing on *who* they are, as the latter is often not critical for immediate response efforts.

In addition to the identified goals and decision process for APT attribution, we observed several challenges in practice. These primarily stem from existing tools' inability to handle the diverse and complex data required for accurate attribution and difficulties in attributing APTs that use standard system tools, shared infrastructure, and overlapping malware. Further lack of standardization in naming conventions affects the merging and correlation of threat information from disparate sources. We also noted issues with internal and external collaboration, which is essential for the more advanced levels of attribution.

In summary, we make the following contributions:

- We offer a comprehensive understanding of security practitioners' processes for investigating APTs, identifying attribution as a layered process that balances the accurate identification of threat actors with the practical considerations of incident mitigation.
- We highlight the various tools and processes used to investigate the three distinct levels, i.e., (1) APT Classification, (2) TTP Attribution and (3) Country Attribution.
- We provide insights into the challenges encountered with these tools and processes, offering recommendations for improving attribution-based research.
- Based on our results, we provide four recommendations for improving attribution tool development and threat intelligence sharing.

Artifacts. We provide the full screening survey, interview questions, and codebook at `https://osf.io/hjdk2/`.

2 Background & Related Work

APT incident response involves a focused set of activities within security operations, including the detection of sophisticated malicious activities, in-depth analysis of attack artifacts, and the attribution of these activities to specific threat groups.

APT Detection and Attribution. Existing APT detection research primarily uses alert correlation to identify anomalous behaviors or APT footprints. For instance, Ghafir et al. [20] developed MLAPT, a machine-learning based system for detecting APTs via network traffic data, while Sachinananda et al. [59] correlated alerts from Intrusion Detection System (IDS), Endpoint Detection and Response (EDR), and Security Information and Event Management (SIEM) systems to cluster those related to the same APT attack scenario. Provenance graphs have also emerged as a state-of-the-art approach in APT detection with tools like ANUBIS [5], APTHunter [40], Unicorn [26], NODLINK [36], and MAGIC [30] using audit logs for anomaly detection.

Prior work on APT attribution has utilized various methods to link malware to specific threat groups. Marquis-Boire et al. [42] manually extracted static features like command and control (C&C) infrastructure to associate executables with their authors. Rosenberg et al. [58] and Wang et al. [69] applied machine learning to classify APT groups using features from sandbox reports and string and code features, respectively, while Han et al. [25] used dynamic API sequences for detection and attribution. Mirzaei et al. [50] identified unknown APT samples through code reuse analysis. Ren et al. [56] proposed a knowledge graph model for attribution using Open-Source Cyber Threat Intelligence (OSCTI). Most recently, ADAPT [60] utilized static features extracted from heterogeneous file types for attribution by clustering executables and documents in threat groups and campaigns.

Expert Studies in Security Operations. In addition to developing technical solutions, human-centered studies have been conducted to understand the cognitive process of software reverse engineering [7, 41, 67, 74]. Votipka et al. [67]

performed a study with reverse engineers in 2020. They developed workflows that represent the necessary process reverse engineers follow and suggest guidelines for designing future reverse engineering tools. This early research has informed subsequent studies that further explore this area [41], and investigate other related fields such as malware analysis. In 2021, Yong et al. [74] conducted a user study specifically to understand the objectives and workflows of malware analysts in practice. Aonzo et al. [7] compared the procedures followed by humans and machines to classify unknown programs as benign or malicious, aiming to understand how data from malware analysis reports is used to reach a decision. They accomplish this by designing an online game that requests participants to classify suspicious files based on their sandbox reports. Prior work has also investigated the usability of tools used by reverse engineers [43, 73] and compared the vulnerability discovering process of software testers and ethical hackers [68]. More recently Maxam et al. [44] and Badva et al. [8] have focused on the broader practice of threat hunting and the associated challenges, with their studies examining the overall process of detecting and responding to threats.

There is a rich line of research on SOC workflows, exploring the general challenges analysts face. Studies have interviewed SOC analysts to investigate their views on security misconfigurations [18], strategies for analyzing sophisticated malware attacks [2], burnout among SOC personnel [63], collaboration between people and tools [21], and the problem of excessive and false security alerts [3, 32]. Oesch et al. [52] examined the usability of two machine-learning based network security tools, identifying issues such as poor documentation and inconsistent UI design, based on surveys of six US Naval SOC analysts. Mink et al. [49] expanded on this by exploring the unique challenges of machine-learning based tools in SOCs, particularly focusing on their explainability and how they differ from traditional tools.

There is also a substantial body of work examining the properties of open threat intelligence (OTI), also known as abuse feeds and blocklists [34, 47]. These studies consistently highlight issues with coverage, timeliness, and accuracy. Recent efforts to measure threat intelligence (TI) quality have primarily focused on OTI, as seen in studies by Li et al. [37] and Griffioen et al. [23]. David Bianco [10] found that contextualized, high-level TI can help address false positives; however, a 2019 SANS survey [14] revealed that respondents still value low-level indicators of compromise more than high-level TTPs. Bouwman et al. [13] explored paid threat intelligence (PTI) and found that experts favor PTI due to its curated and aggregated information. Tounsi et al. [64] demonstrated that (1) fast sharing of TI alone is insufficient to prevent targeted attacks, (2) trust is crucial for effective TI sharing between organizations, (3) standardized TI formats enhance data quality and support better-automated analytics, and (4) the best TI tool depends on an organization's goals, balancing standardization, automation, and speed requirements.

While previous studies have primarily focused on general malware analysis and reverse engineering workflows, user experiences with various security tools, and broader SOC and threat-hunting workflows, our research is specifically concentrated on analyzing the specific workflows, processes, and challenges involved with APT incidents.

3 Methodology

We employ a semi-structured interview protocol designed to get detailed insights into experts' processes. We conducted the interviews with key stakeholders within the SOC [3], including threat intelligence researchers, incident response specialists, security consultants, and malware analysts across various levels of seniority. This approach ensured a comprehensive understanding from multiple perspectives. The interviews were designed to gain deeper insights into how security practitioners investigate and attribute APTs, with a focus on the tools and techniques they employ. Our research is specifically centered on examining the workflows, processes, and challenges associated with managing real-life APT incidents.

Our study consists of two parts (see Figure 1): a screening survey to select qualified participants, and a one-hour semi-structured interview, during which we recorded video and audio, which we later transcribed. Our study was reviewed and approved by our institutions' ethics review boards (details provided in our ethics and open science statement).

In the following, we describe our recruitment, screening survey, interview, and data analysis procedures, as well as our survey's limitations.

Recruitment. We recruited participants over a six-month period (November 2023 – April 2024) through multiple social media platforms (e.g., Twitter/X, LinkedIn), and by distributing flyers at targeted in-person industry security conferences. We also reached out to our personal contacts in various organizations who then shared the study information with their security teams. We recruited participants from security companies, managed security service providers, and the security research domain, all of whom have extensive experience in SOC roles and handling APT security incidents, which was verified through the screening survey. In total, 15 qualified participants completed the interview. Our sample size is sufficient to provide strong guidance for future quantitative work and develop generalizable recommendations for design based on qualitative best practice [24]. We stopped recruitment when we observed that no new concepts or themes appeared from the interviews (i.e., thematic saturation [16]).

Eligibility. We invited participants, who were older than 18 with at least one year of professional experience in dealing with APT incidents and attribution. We assess the experience criterion through the screening survey, where participants self-reported their relevant industry experience and the primary goals of their threat or malware analysis work.

Screening Survey (Figure 1.A-C). Participants began by completing a brief screening survey in which they self-report their job titles, roles, and industry sectors. They further report their expertise in specific areas—such as malware analysis, APT tracking, threat intelligence, and incident response—using Likert scales with options ranging between "Novice," "Intermediate," "Advanced" and "Expert" to capture self-reported proficiency (see Appendix A and our artifact).

Semi-Structured Interview (Figure 1.D-F). We invited eligible participants to a 60-minute online interview, conducted in English. The interview procedure was designed to provide a comprehensive understanding of participants' approaches to handling APTs. Initially, we explored participants' understanding of APTs and the significance of attribution. We focused on their knowledge and experience with APT incidents, including the processes and pipelines used for investigation. To capture a broad range of perspectives, we did not impose a strict definition of attribution; instead, we utilized participant-provided definitions throughout the interviews. This approach facilitated discussions on the varying levels of attribution, including steps taken to identify the type of attack, the attackers, and their tactics (modus operandi). Subsequently, we examined current practices by inquiring about the specific tools and processes employed during APT investigations. This included methods for threat correlation, usage of machine learning, and the integration of Cyber Threat Intelligence (CTI) into investigations and attribution. Finally, we discussed the challenges participants face when managing APT incidents and performing attribution, aiming to identify common obstacles and areas for potential improvement in current practices. Our interview questions are listed in Appendix B and available as part of our artifact.

To maintain consistency between interviews, the interviewer followed a detailed guide on best practices, including how to begin and end the interview, ask questions in a non-leading manner, re-obtain consent, and allow time for participant questions (adapted from Rader et al. [54]). To ensure the clarity of questions and the use of appropriate terminology, we co-designed the interview questions with a usability security expert with nearly a decade of experience and a security threat analyst from the authors' personal contacts. To ensure the questions were easily understandable, we conducted three pilot interviews. The pilot participants were selected to reflect the professional experience and expertise of our target population. The first participant is a security analyst with over 9 years of experience in nation-state threats and AI/ML in security tools. The second participant is a geopolitical intelligence analyst with 12 years of experience in ransomware, cryptocurrency, and the dark web. The third participant is an Associate Professor and founder of a startup, with over 10 years in advanced malware research and cybersecurity education. Following the pilot interviews, we made minor adjustments to the questionnaire and the major themes. Hence, we do not include the pilot interviews in the final data. Our changes included consolidating questions about machine learning usage into the broader theme of tools and processes and added questions about gathering intelligence. Additionally, we incorporated prompting examples drawn from our pilots' responses to help participants better understand the context of the questions if the participant appeared confused.

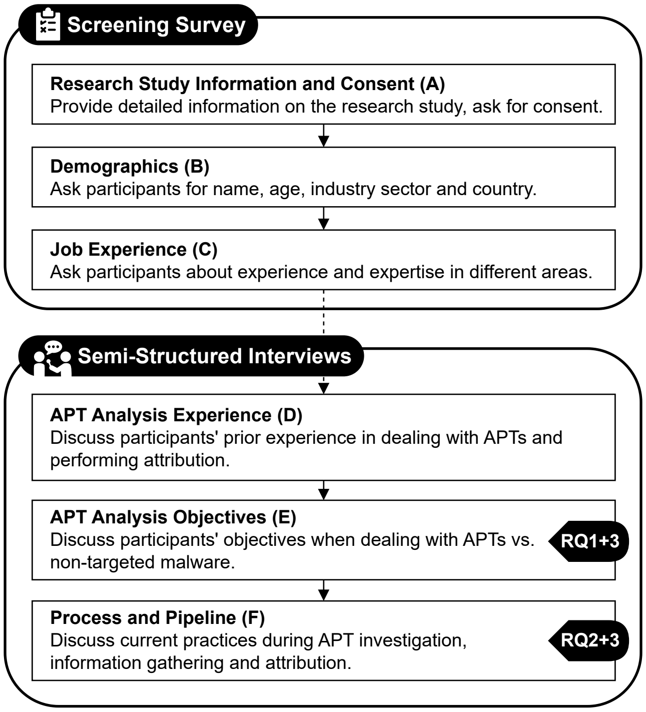
Screening Survey
Research Study Information and Consent (A)
Provide detailed information on the research study, ask for consent.
Demographics (B)
Ask participants for name, age, industry sector and country.
Job Experience (C)
Ask participants about experience and expertise in different areas.
Semi-Structured Interviews
APT Analysis Experience (D)
Discuss participants' prior experience in dealing with APTs and performing attribution.
APT Analysis Objectives (E)
Discuss participants' objectives when dealing with APTs vs. non-targeted malware.
RQ1+3
Process and Pipeline (F)
Discuss current practices during APT investigation, information gathering and attribution.
RQ2+3

Figure 1: Study protocol diagram outlining key stages: (A) Research Study Information and Consent, (B) Demographics, (C) Job Experience, (D) APT Analysis Experience, (E) APT Analysis Objectives, and (F) Processes and Pipeline.

Data Analysis. We transcribed all interviews using the GDPR-compliant transcription service MAXQDA [45]. We then analyzed these transcripts following an inductive thematic coding approach [16]. To establish an initial codebook, two authors collaboratively analyzed two interviews, allowing codes to emerge from the data and then discussing the initial codebook with the full research team. The two authors then independently coded interviews in rounds of two. After each round, inter-rater reliability (IRR) was calculated using Krippendorff's Alpha (α) to account for chance agreement during coding [28]. Then, coders met to resolve disagreements, change the codebook as necessary, and apply changes to previously coded interviews. The full research team met to discuss the results after each round and to review proposed changes to the codebook. These changes included identifying and merging overlapping codes, as well as adding new codes to reflect emerging themes, such as future improvements that our interviewees suggested as the field of APT attribution progresses. After four rounds of independent coding (eight interviews), an IRR of $\alpha > 0.8$ was reached for all subjective codes, indicating high agreement [28, 39]. We did not calculate alpha for objective variables like tools mentioned by participants, as these can be inferred directly from the transcript. The five remaining interviews were coded independently by one author. The final codebook and α values for each variable are available as part of our artifact.

In the next phase, we performed axial coding to explore the relationships between and within these categories [16]. We aimed to develop a theoretical model by extracting and organizing themes from the coded data. We identified three primary categories related to the handling of APT incidents i.e., attribution value, attack analysis (including processes and tools), and the challenges encountered within each process. We further linked these challenges to participants' future recommendations and suggestions. From these connections and relationships, we derived a theory that identifies the high-level processes and specific technical approaches used by analysts.

Limitations. As a semi-structured interview, some follow-up questions may not have been asked in every session, and participants' responses might not cover all topics with the same depth. This is especially common for expert tasks [6]. We produced a thorough script and a single interviewer conducted each interview to improve consistency. None of the participants spoke off the record. It is important to acknowledge that due to the limited amount of time per interview, certain themes might not be covered in participant's responses. This further adds to the motivation to not generalize the findings based on the frequency of specific responses.

Second, there may exist concerns that the participants do not fully represent the entire population of security professionals who deal with APT-related incidents, such as government officials. Although we recruited participants from diverse professional and demographic backgrounds (see Table 1), our sample is predominantly centered on US and EU participants. Moreover, it does not comprehensively represent all possible roles, industries, or demographics. To address this limitation, we ensure careful interpretation of our qualitative results and do not attempt to generalize our findings. Instead, we focus on capturing a diverse range of perspectives from various stakeholders within the community. These findings can serve as a foundation for hypotheses in large-scale surveys or targeted studies focused on specific professional or demographic factors in future research. Finally, biases such as social desirability and confirmation bias may influence some participants' responses. We mitigated these by framing questions in a neutral manner and encouraging participants to consider and discuss opposing viewpoints.

Table 1: Participants in our study along with their roles, organizations, and years of experience. *Role key:* MA = Malware Analyst, IR = Incident Response, TI = Threat Intelligence. *Organization (Org) key:* MSS = Managed Security Services, SC = Security Company, SR = Security Research, IntSec = Internal Security Team (Tech, Financial, Healthcare). *Size key:* S = Small, M = Medium, L = Large.

ID	Job Title (Experience in Years)	Sector	Org Type	Org Size
P1IR/MA	Research Director (22)	Industry	MSS	M
P2TI	Managed Defense Head (13)	Industry	MSS	M
P3IR/TI	Security Consultant (6)	Industry	SC	S
P4IR	Research Scientist (12)	Non Profit	SR	S
P5MA	Senior Malware Analyst (18)	Industry	IntSec	L
P6TI	Threat Intelligence Researcher (5)	Industry	SC	L
P7IR	Security Analyst (10)	Government	SR	M
P8IR/MA	Security Researcher (10)	Industry	SC	M
P9IR	Security Operation Lead (14)	Industry	MSS	M
P10TI	Manager CTI (15)	Industry	MSS	L
P11IR	Security Engineer (17)	Industry	IntSec	L
P12IR	Sec. Operations Director (15)	Industry	SC	L
P13IR	Senior Threat Hunt Analyst (9)	Industry	SC	L
P14IR/TI	Threat Operations Lead (16)	Government	IntSec	L
P15MA/TI	Sec. Consult. Manager (10)	Industry	MSS	L

4 Participants

We had 15 participants who completed the interview. All participants had more than five years of experience, with most having over ten years, specifically in SOC operations. By interviewing participants with extensive experience working at well-established security groups in large tech companies, leading security industry organizations, and government agencies, we were able to identify a wide range of perspectives and gain a comprehensive understanding of how APT incidents are managed in practice. Table 1 shows the list of all participants, their job title, sector, type of organization, size of organization, and years of experience. Professionally, our participants consisted of a variety of roles, including first-level responders to active security alerts, upper management in their organizations, senior malware analysts, research scientists, and threat intelligence researchers. Five participants worked for managed security services, five for security companies, two in security research and advocacy, and three on tech, financial, or healthcare institutions' security teams. Eight participants were from large organizations with more than 5,000 employees (several exceeding 100,000), five from medium-scale organizations with 50-5,000 employees, and two from small organizations with fewer than 50 employees. We provide this information about participants' roles and organizations only to add context and demonstrate the sample's diversity.

Our participants reside in a variety of countries, such as the UK, USA, Austria, Canada, Israel, and Finland. On average, participants spent about 70 minutes completing both the survey and the interview (60 minutes of which were the interview). The majority of study participants identified as men; two identified as women. Our participants were educated (i.e., all had a Bachelor's degree and eight had a graduate degree). Additionally, our participants reported having an advanced level of skill in at least one area relevant to attribution, i.e., malware analysis, APT attribution, threat intelligence research, or incident response.

5 Result: Goals of Attribution (RQ1)

In this section, we discuss the importance of attribution and explore scenarios in which participants expressed interest and reasons for attributing an incident. Note, through our interviews, we do not attempt to generalize the prevalence of specific practices across all APT investigations as certain practices may not be applicable in all contexts or roles. Instead, we enumerate the range of practices and tools present generally in APT analysis to support future quantitative investigation. To enrich our findings, we include incident and organizational detail whenever participants provided them. However, it is important to note that participants shared experiences from incidents they had handled throughout their careers, sometimes referencing past organizations they worked for, without offering detailed descriptions of specific incidents or organizations. Additionally, we categorized roles based on participants' job titles and responsibilities. We observed from our interviewed samples that APT attribution often involves multiple roles such as malware analysts, incident responders, and threat intelligence analysts, sometimes filled by the same person. However, we found that the tasks associated with these roles were not always consistent across different organizations. There are similar indications in prior research [12] regarding the correlation between job titles and the tasks performed, suggesting that while job titles may guide expectations, the actual tasks can vary based on other contextual factors. In our study, we observed that in-depth reverse engineering of malware was exclusively handled by senior malware analysts. Apart from this, we did not observe clear differences in themes across the reported roles, suggesting that our findings apply broadly across different aspects of the work.

TTP attribution informs investigation and effective threat prioritization. Several participants (N=8) pointed out that some level of threat actor attribution is useful for understanding the threat and guiding incident response. P3IR/TI noted (TTP) attribution helps in, "understanding what TTPs to look for in their network." P11IR further explained being able to "attribute a binary" to a "specific threat actor" helps in pulling other "indicators or TTPs" that are known for that specific threat actor and use that as a way to perform a deep dive investigation, further adding that TTP-level attribution "provides a lot of pivot points to be able to search for other things" in the environment and being comprehensive in the investigation. Participants emphasized that even partial attribution, such as classifying a threat as generic versus an APT, streamlines the incident response protocol (N=3). This classification allows organizations to effectively deploy specialized forensics teams, trigger adherence to specific protocols or engagement with law enforcement, and expedite remediation when necessary. Understanding whether an attack is specifically targeted or a "spray-and-pray" approach helps organizations to prioritize resources on addressing high-risk, targeted threats while de-prioritizing more generic, less impactful attacks. P6TI explained "If we don't know what the threat is, we don't know how severe or how to prioritize the attack. So [attribution] is quite a good way to know whether you're being targeted or whether it's sort of spray and pray." P10TI further highlights the importance of accurate attribution in ransomware attacks. They added understanding the adversary and "what assets you have that are so attractive" allows for well-informed remediation when you "need to negotiate" with the attacker.

Balancing incident mitigation with full attribution for strategic decision-making. Participants emphasized (N=7) that victim organizations' primary initial concerns are not identifying the specific perpetrator—what we will refer to as country attribution going forward—, but rather understanding an incident's full scope. The question of who carried out the attack is often secondary, unless the target is politically sensitive, where country attribution carries more significance. For example, P1IR/MA noted "This was China, or this was Russia or this hacking group; that aspect comes right at the end, if at all. We're not really too bothered if we get to that point or not because... [the client] just want a warm, fuzzy feeling that [the attackers] are out of the network." However, participants (N=4) mentioned cases where full country attribution becomes valuable particularly when understanding the origin of an attack can significantly impact response strategies. For example, P6TI described their thinking assuming a scenario where they were operating a Ukrainian network and detected lateral movement saying "If it's Russia, they're going to go straight for the domain controller and then wipe everything. If it's China, they might try and persist in there a little bit longer... you would want to try and stop the Russians first because they're going to destroy the whole network."

In such scenarios, prioritizing responses based on the likely actions of the threat actor, such as stopping Russian actors who may destroy the network versus Chinese actors who might engage in prolonged data exfiltration, can be critical for effective incident management. While participants highlighted the importance of rapid and accurate full-country attribution to deploy strategic and timely countermeasures, our study does not delve into the specific details of the mitigation process, as this was not the central focus of our discussions. Instead, we analyzed the goals and technical aspects of attribution and how it informs incident response.

Further, country attribution is necessary in cases when an incident might have geopolitical, legal, or strategic business consequences. As P4IR explained, precise country attribution can be used "to make policy, to respond, you know, and different entities can respond in different ways, right? Like a company might want to understand attribution, like, you know, Google, like, oh, our our servers are being hacked by a Chinese group. We're going to withdraw Google from China, right? Like they did in 2009, 2010." When participants described this level of attribution, they were quick to point out that a high level of caution and confidence is necessary as misattribution can cause geopolitical tensions or legal challenges and risks escalating conflicts. P3IR/TI remarked "You have to be extremely cautious when you're saying something like that. Right. And that level of attribution, because there are greater implications."

Country attribution helps with long-term remediation and proactive measures. While they might not care about full country attribution during the incident response, a few participants (N=3) recognized its importance for informed, long-term organizational strategy. P6TI explained "If you know a specific country, countries, APT groups are coming after you, then you know you should focus your research on the capabilities of those APT groups. . . you know, trying to be kind of predictive in a way."

6 Result: Processes and Tools (RQ2)

In accordance with the responses described in Section 5, we observed a decision tree describing how participants moved through increasing levels of attribution specificity. We begin this section by summarizing the decision tree and providing a visual depiction in Figure 2. Progression through the decision tree depends on the characteristics of the identified threat and the participants' organizations' motivation. As participants described moving through the decision tree, additional features were considered, and analysis processes were conducted to provide more detail. We describe the specific features and processes for each level in turn in this section.

Attribution Decision Tree. When handling a potential APT-level incident, the first step is to determine whether it qualifies as an APT. Participants reported certain characteristics that help distinguish an APT-level attack, as discussed in Section 6.1. If the incident does not meet these criteria, standard triage procedures are followed to contain and mitigate the threat. However, if the incident is confirmed as an APT and the organization has processes for TTP attribution, it triggers an advanced response involving different teams and an in-depth forensic process, as discussed in Section 6.2. The investigation then focuses on the dedicated remediation process by identifying various attributes and TTPs associated with the incident mentioned in Section 6.3. If the organization aims to identify and attribute the incident to a specific country, it further gathers and connects additional intelligence to pinpoint the entity behind the incident discussed in Section 6.4.

6.1 APT Classification

Our participants reported that the first critical step toward potential full attribution is determining whether it is, in fact, an APT (see Figure 2.A). To this end, participants considered several indicators of potential APT activity. We discuss each below. Note that this layer of partial attribution acts as a classifier for initial triage, so any of these indicators can be sufficient to signal an APT.

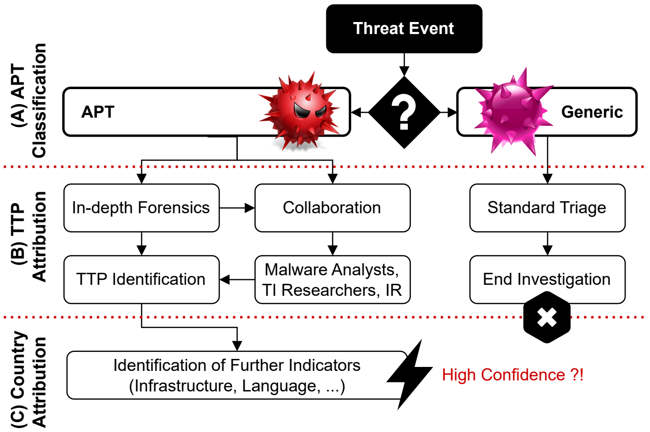

Figure 2: Decision tree outlining the process for attributing threats, starting from *APT Classification*, followed by in-depth forensic analysis and *TTP Attribution*. The decision process ends with a high level of confidence at *Country Attribution*.

APT attacks are characterized by low-profile tactics and targeted efforts. Participants (N=7) emphasized that APT actors rely on stealthy techniques, such as 'living-off-the-land' methods. By using existing tools and legitimate system processes, these actors blend seamlessly with regular activity, making detection more difficult. P9IR noted, "The attack is not noisy at all. It's very targeted. . . that is going to be a problem for any automation that you have, finding anomalies like that, because they are not an anomaly." P6TI suggested this targeted nature makes it necessary to compare information from other organizations to assess whether an attack was specifically targeted. They explained, "Our email gateway that we use, we can actually see all the other customers that also receive the same email as you. So that's quite a good way to know whether you're being targeted or whether it's sort of spray and pray. Because if a thousand other customers also receive the same email that you did, then it's not really going to be targeted." This method, they added, helps prioritize threats that require more in-depth threat hunting and due diligence.

APTs are characterized by their use of lateral movement. Participants (N=7) emphasized lateral movement as a key indicator for targeted attacks. APT actors are highly skilled in conducting extensive reconnaissance over several months to identify critical targets and initiate lateral movement. P9IR noted, "It's the work smarter, not harder mentality that drives the most success for APT actors." Participants observed that APT actors spend considerable time understanding and blending into systems, making them difficult to detect. As a result, lateral movement often serves as the first sign of their presence, as other actions tend to be slow and subtle. P9IR elaborated, "In many cases, when we catch them, we find that six or eight months earlier, they were already conducting reconnaissance in that space." P12IR further explained the importance of tracking APT actors' lateral movement, as it provides additional insight into their goals and the broader scope of the attack, mentioning, “This group over here is interested in point-of-sale machines. How do you know? Because they tried to attack it. They didn’t get there, but they tried to move laterally to point-of-sale systems. Now we know that’s part of their objective, even after containment.”

APT attacks are characterized by multi-stage components. Participants (N=4) noted that APTs are often defined by their multi-stage, coordinated nature, with different components playing distinct roles in executing various parts of the attack over time. As P13IR explained, “Incidents often don’t happen with just one person doing things. It’s usually a continuous chain, where there might be malware on a system from a while ago that stole credentials or something like that. Then, the actors hand it over to someone else to take action on the box and do something against it.” This highlights the importance of connecting seemingly disparate events and recognizing them as part of a larger, orchestrated attack rather than isolated incidents. P7IR stressed the need to thoroughly scrutinize each component in APT attacks, noting, “If you’re dealing with cheap, off-the-mill malware, and you see it leaving a file, you might think it’s just an error or a leftover artifact.” But when dealing with APTs, they recognize different components as being part of a deliberate and coordinated effort and “look at it differently because [advanced actor] wouldn’t just leave a file lying around randomly.”

6.2 Incident Response Process for APTs

If an incident is classified as an APT-level attack, i.e., targeted, novel, and more sophisticated, participants (N=8) reported following a dedicated, more comprehensive, and collaborative protocol. This is consistent with Wong et al.’s malware analysis workflow, which suggested analysts perform more in-depth reviews when working with novel malware [74]. Below we provide further detail on the in-depth incident response (IR) process. Note that this process applies to both levels of attribution (TTP and country) as identified by the first layer in Figure 2.B. Additionally, we did not observe any clear differences in processes across participants from different organizations (industry, government, non-profit), apart from variations in the type of APT actors they prioritize. For example, the non-profit organization focused on analyzing and addressing threat actors targeting civil society groups. Meanwhile, the majority of our participants are from industry, and discussed prioritizing threat actors likely to impact their high-risk assets like cloud resources. Given our small sample space and lack of comprehensive data, these observations cannot be generalized to each industry, however, they may inform future research.

Analysts perform in-depth analysis for cases involving sophisticated or novel threats. Participants’ response strategies shift from internal management of lower-level threats to increased investigative efforts for APTs. Participants noted that handling sophisticated threats involves a complex and layered approach (N=8). Initially, standard incident response procedures are followed, including memory dumps, disk images, log analysis, and network traffic examination. However, as P13IR noted, when an investigation reveals more serious indicators, such as attempts to manipulate code, the response escalates, and they “spin up a big bridge” and “get a bunch of other people… to take a look at [incident] and see what happened.” This escalation often involves mobilizing a broader range of experts across the organization and conducting a detailed forensic investigation. For example, P11IR described doing a more extensive investigation of further lateral movement by examining “every single network artifact that has been touched,” in contrast to routine malware responses that might only involve rolling credentials, disabling access, wiping, and reformatting devices.

APT incidents involve cross-team collaboration and coordination with external entities. P5MA, who is from a large organization (>5,000 employees) with a clear distinction between security teams, describes the structured, collaborative nature of an APT investigation, highlighting the division of labor between teams. In this scenario, “the malware team dissects the threat,” while “threat intelligence is connecting the dots.” P6TI further elaborated on the role of the threat intelligence team in guiding investigations and escalation. They explained “If we [the CTI team] do have even low confidence that an APT group is actually targeting us, we’ll take it more seriously.” They elaborate based on the initial bit of information they decide whether a threat is worth “becoming an investigation.” Meanwhile, P7IR, who is from a government organization where privacy protocols differ significantly from those in the private sector, emphasized that “certain privacy precautions can be disabled and additional documentation is required” to ensure proper handling and coordination with external entities, such as other government agencies.

6.3 TTP Attribution

In this section, we discuss how, as part of the detailed investigation and collaboration, certain key activities are undertaken for TTP attribution (see Figure 2.B). Participants use historical knowledge to correlate incidents, conduct detailed analyses of malicious artifacts, and gather trusted intelligence. The ultimate goal of these steps is to identify the actions and objectives performed by the threat actor by closely attributing the tactics and techniques used to known patterns, thereby mitigating the incident’s impact.

Participants use historical knowledge for correlating incidents. Participants (N=13) emphasized the importance of historical knowledge and threat intelligence in analyzing and correlating APT incidents. This process involves correlating and comparing the interconnected attack chain components—common in APTs (see Section 6.1)—to historical records, such as threat data from prior incident response engagements. As P1IR/MA noted “We leverage our research to collate and understand commonalities across engagements, building our own knowledge pool of adversaries." They further explain that they use automated methods for IoC hunting across the organization, stating "We focus more on TTPs and IoCs, maintaining our own internal databases of confirmed threats." Participants also mentioned using manual methods for the historical lookups with P6TI using spreadsheets "with the date [threat report] was published, the source, the adversary [involved]" and P9IR documenting low-confidence indicators for future reference explaining to "note those things" to see if they have seen them before. P14IR/TI highlighted the maturity of their approach by emphasizing the focus on "TTPs rather than IoCs." Based on the concept of the Pyramid of Pain [10], they explain that "hunting based on tactics" is more robust, as relying on IoCs alone can generate an overwhelming number of "alerts."

Participants perform malware analysis using initial correlation of IoCs through threat intelligence, followed by in-depth reverse engineering. Most participants (N=11) investigate and correlate malware or file artifacts found in the attack chain using OSINT. This involves querying platforms like VirusTotal [66], Triage [19], and Joe Sandbox [31] to gather information about malware samples. Analysts consult these public forums to identify any existing intelligence that can aid in their investigation. As P11IR noted "Most of the time, we'll just do searches for the hashes to see if it's already been detected in the wild" and highlighted the use of subscription-based threat intelligence and private contracts noting "if we are confident that it is like an APT-level attack, then we would really leverage our internal, the private [threat intelligence] contracts... [to get] all of the details that is associated with the file, and the capabilities of the files."

Additionally, participants mention using a fuzzy hashing approach, such as ImpHash [55], peHash [71], SSDEEP [33] or internally developed techniques, to identify malware from the same family. For instance, ImpHash (Import Hash) generates a hash based on the import table of an executable, focusing on the functions or capabilities of the binary. Analysts typically obtain these hashes from malware analysis tools. These hashes are then used to search for other binaries that exhibit similar behaviors. As P11IR explained, "a lot of times you can then get a hash of those capabilities and then do searches on similar hashes to be able to identify if there are other files that match." P7IR summarized the "two stages of attributing malware." The first stage involves correlating IoCs "by running the malware in a sandbox environment and seeing if it is trying to contact domain X, or it's creating file Y, or it has a request pattern that we've seen with malware Z." The second stage requires a deep dive into reverse engineering with tools like Binary Ninja [11] or IDA Pro [29], looking for "specific techniques, code styles, and decisions in the program logic" that can help correlate with other samples and is often done by dedicated malware analysts. However, P10TI also remarked when it comes to APT incidents while analyzing malware is important, "there's more than the malware." They emphasized the need to consider additional context, such as phishing lures and explained "Let's say the malware was propagated via phishing lures. There's the lures themselves. What do the lures say? Who could the lures be targeted at?" This broader context is important for a comprehensive analysis.

Participants use trusted relationships to share detailed information about APTs that cannot be publicly disclosed. To exchange information about APT incidents and relevant artifacts, participants (N=7) emphasized the importance of trusted circles—a selective group of individuals who share intelligence privately among members, also mentioned in prior work by Bouwman et al. [13]. Sharing information on APT incidents is difficult due to the sensitivity and strategic nature of the data. P4IR discussed the balance between the need for information and the risk of over-disclosure, noting that trust is crucial for private exchanges: "When you're tracking a threat group and the threat group is trying to avoid being tracked, you don't want to give away too much." P6TI, from a medium-sized organization, described the daily sharing of threat data within a trusted circle, which includes discussions such as "Has anyone observed exploits targeting this CVE?" or "Has anyone seen this malware?" They described the operation of a Trust group, saying, "I also run the [anonymized] Trust group... a group of intelligence researchers that share threat data and threat information with each other on a daily basis. We have about 150 analysts." This behind-the-scenes collaboration enables the exchange of sensitive details about malware, CVEs, and IoCs facilitating attribution efforts.

6.4 Country Attribution

Toward the final stages of the investigation, when participants seek to precisely attribute the actor's location or origin, they use a combination of key attributes alongside identified TTPs and IoCs to perform country attribution (see Figure 2.C). Given the sensitivity and difficulty of this task, participants rely on confirming evidence from multiple indicators and look for clues beyond those considered in earlier levels of attribution, such as linguistic patterns or specific wording used by the threat actor.

Participants rely on IP addresses, domains, and C&C infrastructure. Most of the participants primarily rely on infrastructure elements such as IP addresses, domains, and command and control (C&C) channels for country attribution (N=12). P9IR elaborated that using IP and geolocation data can "narrow down the pool to a few possibilities. Then we would search for things like ISPs." P12IR further emphasized the importance of metadata-related infrastructure features, noting that "ASN off the IP" and "any registration data" such as the email address of the registrars, are useful for attribution. Participants also mentioned using time zone analysis to infer the threat actor's hours of operation and potentially their region (N=4). P8IR/MA explained that the North Koreans "were working six days a week from 9 to 9" which provides a useful indication for attribution based on the working hours of attacker and the timeline of attack.

Participants investigate the choice of wording. Participants (N=7) highlight the importance of analyzing the choice of wording and language within communications or malware code to infer the geographic or cultural origins of the threat actor. P1IR/MA points out that seemingly minor details, such as "the choice of the password or passphrase" can be quite revealing. Sometimes the actors "might throw in cheeky comments in their code" that can help in identifying the actor based on their vocabulary and language. P13IR adds that actors may "leave a message in the registry key" that could serve as their "call sign" significantly increasing confidence in identifying the origin of the attackers.

Participants investigate reused tools and exploits. Participants (N=7) also view reused binaries and exploits from previous incidents as valuable indicators for attribution. P1IR/MA mentions that a "[nation-state] state might have reused or slightly modified a piece of command and control" or a "backdoor Trojan" recovered during the incident allowing analysts to "infer some attribution" based on their familiarity with "something similar." P6TI further explains that if a threat actor uses a "certain piece of unique custom malware" it helps in attribution since nation-state actors who have "developed it themselves" and have not "shared it around to anyone else" leave a distinctly identifiable footprint.

Participants build threat actor profiles to inform and guide long-term remediation. One of the goals of country attribution is proactive threat actor tracking to guide organizational strategy (see Section 5). Towards that end, participants highlight the importance of developing comprehensive threat actor profiles for effective threat management. This process involves gathering intelligence from various sources, including OSINT, commercial threat intelligence platforms, incident response data and trusted intelligence-sharing groups. P6TI elaborated on this process within a sizeable CTI team: "We will divide our analysts up into regions like what regions they should be focusing on ... And I'll perform a quarterly report on Russian APT group." P6TI further explained how they extrapolate key information and perform mapping, stating "I'll try to extract the information from [threat] report into the diamond model." The Diamond Model [70] is a framework that examines the interactions between four key elements in a cyber attack: the adversary, their capabilities, the infrastructure used, and the victim. By mapping these relationships, the model allows analysts "to basically build up the database" for long-term planning. In practice, this process involves a deep analysis of threat intelligence sources, such as vendor reports and blogs, to associate aliases with specific threat actors. Analysts then assess how elements of the Diamond Model—such as adversaries, infrastructure, and TTPs—overlap across different reports. As P6TI described, "that's how we kind of get to Diamond Models that basically mean the same threat group." These overlaps help in building accurate threat actor profiles, enabling analysts to confirm that groups identified in various sources are indeed the same.

7 Result: Challenges (RQ3)

In this section, we explore the challenges participants encounter in investigating APT incidents and performing attribution. Section 7.1 highlights issues with relying on infrastructure features and the limitations of file analysis automation as APTs become more sophisticated. We also discuss the difficulties in data ingestion and the minimal use of machine learning, which impacts rapid correlation and attribution. Section 7.2 discusses challenges in current processes, such as the fragmentation of threat information across databases, inconsistencies in naming conventions, and the complexities of different threat intelligence sources, all of which affect the accuracy and integration of threat data.

7.1 Challenges in Tooling

Participants reported a wide range of specific challenges they faced when using tools to perform attribution related tasks. These were most often related to existing tools not supporting the data types and formats necessary for successful attribution as APTs and the TI ecosystem grow in complexity.

Lack of automation and validation in data ingestion impacts the use of historical threat data. As we discussed in Section 6.3, a key process for TTP attribution is querying threat data across historical records of threat intelligence and using a database of IoCs and TTPs. However, multiple participants discussed challenges in collecting and using this data (N=3). For example, P4IR mentioned a lack of tools "for crawling websites that publish reports or ingesting indicators from those reports," indicating a need for more automatic data ingestion and comparison. Participants acknowledge that for malware, there is a little more automation using machine learning; however, malware is just a small "part of a threat actor TTPs" (P10TI). To gain a sufficient understanding of threat actors, participants indicated they had to expend significant manual effort reading everything that has been written about a particular threat actor (N=9). As P13IR explained, maintaining threat actor information involves "a lot of parsing...It's a very manual effort."

Additionally, participants reported difficulty in not only collecting all relevant data but also validating it to avoid false positives (N=2). P13IR gave an example false positive describing the automated processing of threat intelligence reports which will indicate "the domain for this is google.com/something. Then, your intel feed down the line will be like let's look at all the domains in the reports here and then they'll say Google is bad." This is a growing concern as APTs more often use common or public infrastructure (see Section 6.1).

There is a lack of advanced and robust tooling to effectively analyze a variety of file formats. For malware correlation participants (N=2) face challenges when analyzing and correlating a diverse range of file types, particularly with the rise of cross-platform binaries. As an example, P10TI notes the shift from traditional languages like C and C++ to languages such as "Nim, Rust, and Golang," which allow threat actors to "target multiple platforms" simultaneously. Further, P5MA highlights the difficulties of dealing with these advancements, stating "We lack proper tooling to analyze files used in these big chains of attack." They explain "Especially for Windows and specific languages like .NET, Visual Basic, and Delphi ... you need a lot of information, and there is a lack of tools" for effective analysis.

APTs using existing system tools rather than custom malware is making it difficult to attribute their activities. Further, to complicate correlation and attribution efforts, APT operations blend in with legitimate tools as a means of being stealthy (see Section 6.1). P7IR notes the increased use of built-in tools like PowerShell and threat actors pivoting toward "living off the land" (N=4). They explain: "We've observed more instances where attackers avoid using their own binaries or scripts as much as possible, relying instead on tools that are readily available on the victim's system." This poses challenges for traditional binary-based correlation methods as P11IR highlighted "There's been a really big push for some of the bigger APT groups to move away from binaries."

APTs exploit legitimate accounts and activities to evade automated detection systems. Participants highlighted key challenges in detecting and mitigating threats within complex environments (N=4). P15MA/TI mentioned that while automation streamlines a lot of their forensic analysis, such as mapping MITRE ATT&CK TTPs [51] to the specific threat actor, it often requires manual review and additional context as the automated system does not understand the "full context of the incident...if it was actually a legit user using their account or if it was the threat group using it." This challenge is exacerbated by attacks involving dormant adversaries within large networks. P6TI explained "[adversaries] can linger in your environment for years at a time. They only need to create an account and just keep it there...if you have an Active Directory with tens of thousands of users, it's really difficult to go through all of those and check."

Application of machine learning in APT correlation and attribution is limited. Despite the recognized need for robust and advanced automation for detecting and correlating APT activities, participants expressed reservations about utilizing machine learning (N=5). Key barriers to broader adoption included a lack of training resources, time constraints, insufficient datasets, the complexity of the models, and high false positive rates in alert generation. P7IR noted, "We have experimented with [machine learning]. But the results have not been, I don't want to say they have been bad. But not worth the effort." They further explained that this was not "necessarily a critique of machine learning approaches, it's more the reality of being severely resource constrained." P7IR explained this resource constraint was on the time available to tune machine learning tools to their specific environment, saying, "The only reason why we even experimented with it in the first place was because I decided to not sleep one night. And I couldn't justify, spending more of my own time on it because during regular office hours, I had other tasks to do." Beyond time constraints, we identify challenges with insufficient resources to train and fine-tune the models. P9IR added to this, emphasizing the challenges in "finding the resources to be able to train the models to do what you need," and pointed out that if not done properly can lead to high false positive rates and "ticket fatigue" in SOCs. These results are similar to Mink et al.'s [49] findings in discussions with SOC analysts regarding their use of machine learning for intrusion detection. Finally, another challenge in ML adoption lies in the lack of diverse malware datasets and the difficulty in explaining the decisions of complex models. However, the data scarcity problem for attribution is further complicated by the fact that effective machine-learning-based attribution requires many different comprehensive datasets. As P5MA explained, "To have the data set with all the functions for the different architectures for the different operating systems...it's really complex."

Apart from learning-based models, two participants mentioned LLMs for APT investigations. One described using internal GPT models for text summarization, noting that this usage is ad-hoc and not a company-wide system. Another participant shared their team's experience experimenting with Microsoft Copilot, highlighting its potential but also issues in dealing with meaningless and incorrect results. It is important to note that at the time of the interviews, LLMs were gaining traction, so it is possible that there have been changes in their use since then. Even so, our results offer insights into the process and can guide their effective adoption.

The reliance on IP addresses and domain names for country attribution is unreliable. As we discussed in Section 6.4, participants rely on infrastructure features such as IP addresses, C&C infrastructure, and domain names to identify threat group signatures. However, P11IR noted that "IP addresses and domains are not nearly as reliable of a correlation point anymore," highlighting the difficulties of APTs blending with legitimate infrastructure or using shared public infrastructure (see Section 6.1). P8IR/MA further points out that attribution becomes more complex in cloud-native environments, where "containers and instances are popping up and down all the time," making it challenging to track persistent infrastructure as it has become super easy to "spawn on a new machine somewhere in the world and just attack with it." Therefore, in most cases, this makes country attribution impossible in practice based on these infrastructure features.

APTs using shared infrastructure, overlapping malware, and selling attacks further complicate country attribution.

Participants highlighted scenarios where attribution could be misleading or faked with P4IR stating "I do have some experiences of attribution getting very murky, like cases where, you know one threat actor might compromise and use another threat actor's infrastructure. Like we've seen some potential cases or indications where it looks like that might be what's happening" which could lead to "all kinds of misattributions." P6TI mentioned another scenario where "one company was developing all the Chinese malware that like ten different Chinese APT groups were using. So it's kind of a this is it comes back to this thing of, we may know it's Chinese APT group or China based group, but we don't know exactly which one," because they all share many capabilities these days. Finally, P9IR noted another challenge unique to country attribution is that "A lot of the APT cases had teams where we could see the A team that is doing the attack, and then the B team doing the attack. And usually, the B team is how we find them. But there's also once they're done with the attack, we know that they sell their attack on the dark web. And then criminals, just regular criminals could then use it in their attacks." These scenarios make it increasingly difficult to accurately identify the true source of APT attacks.

7.2 Challenges in Processes

In addition to the challenges our participants faced when using tools, they also encountered challenges in establishing an accessible and reliable threat intelligence ecosystem, as well as with effective collaboration within and among organizations.

Inconsistent data formats and naming conventions add difficulties in merging and correlating threat information from disparate sources. Participants highlighted the challenges of lack of standardization in threat information across different databases, even for government organizations (N=3). P9IR noted "CISA has a database for tracking one set of threat events, while the FDA maintains another for different events." This fragmentation complicates the process, as both organizations might be tracking the same actor without realizing it. P7IR further emphasized the challenge of sharing information as it might reveal sensitive data. They explained "No government is sharing their attacks with other governments... There are some standards like STIX, or using MISP, but in practice there is no secure way to do this."

Inconsistent threat naming conventions further complicate the process. Multiple participants noted that varying names for the same threat actors between reporting organizations adds to the confusion (N=13). P1IR/MA attributed this to the fact that they "haven't necessarily gone through a due diligence process to see what's out there already or made sure that this makes sense in relation to other publications that are similar... everyone's just busy speaking out what they think is useful... I think that's probably a large part resulting in some of the ambiguity and misinformation that we see."

Open-source threat intelligence has too many false positives; commercial products are too slow, so participants turn to unofficial sources. Participants in our study use both commercial threat intelligence and OSINT in their processes for TTP and country attribution (see Section 6). In addition to tooling performing poorly when attempting to ingest diverse TI, as described previously (see Section 7.1), participants identified issues in threat intelligence itself (N=7). Bouwman et al. [13] and Li et al. [37] already highlighted significant gaps in the accuracy, coverage, and timeliness of threat intelligence sources. Our results suggest a similar set of perceptions. First, our participants found OSINT unreliable due to its lack of important details and context. P2TI explained they have "tons of information which means you have more quantity and less quality. Less quality threat information means it's absolutely not attributed... What can I do now with the list of IP addresses?" The lack of quality information places the burden on analysts to validate the data—specifically, to understand which IP addresses are malicious, how long they should be blocked, and how to properly integrate this information into defense technologies.

Paid threat intelligence also has its challenges. When discussing paid threat intelligence reports, P14IR/TI said they were "probably our slowest means of actual detection creation. They tend to be quite granular but are less actionable." P14IR/TI elaborated that Mandiant or similar services "might identify a malicious IP address from an attack at another organization and include it in their intelligence product. By the time it reaches me as a customer, the attack has likely moved on." P9IR also echoes similar timeliness issues with the IoCs in commercial threat intelligence: "You have to be good about vetting the information" to make sure that the "IPs and domains are still valid, and the attacks are still relevant." In addition to supporting prior results, we discovered our participants go beyond official threat intelligence to unofficial sources like Twitter/X, Reddit, GitHub, and blog posts (N=3). P14IR/TI explained that they supplemented official threat intelligence through their own "live analysis of samples that are on the Internet... Twitter is a great source for this. If someone's seeing something that we think is malicious, then we can then take that in and stuff like, and use our own telemetry to work out what's going on."

Need for collaboration and open communication among teams for successful APT investigations. Several participants emphasized the role of cooperation and transparency among different entities for successful APT investigations (N=5). P9IR remarked "In all the different places I've worked, the tools have varied. For me, it's really about the people and having a collaborative team that is skilled and knowledgeable." They further noted that in siloed environments they "rarely saw progress toward identifying the source or attribution of the attacker" with investigations often hitting a wall. P9IR specifically highlighted an information sharing barrier when people "would only share details on a need-to-know basis" delaying the investigation. P3IR/TI expressed hope for a "shift in the industry" towards "more open and honest reporting on the activities of APT groups." P7IR echoed this sentiment, saying "Cooperating together is the only chance we really have at being successful."

8 Discussion and Conclusion

Attribution is a complex and nuanced process that balances the need for accurate identification of threat actors with the practical considerations of incident mitigation. Our findings highlight a disconnect between theories and practical realities, identifying that victim organization progresses through three distinct layers of increasing classification specificity depending on the incident and their situation. This decision process suggests that attribution should not be viewed as a single task but rather considered from one of the three layers identified, i.e., (A) APT classification, (B) TTP attribution, and (C) country attribution. Each layer presents unique goals, accuracy requirements, and challenges. Further, we found our participants often prioritized incident mitigation over identifying the perpetrator, focusing on understanding the incident's scope, assessing data compromise, and securing networks (see Section 5). Future research should explore whether this motivation extends beyond our sample, as our study offers valuable direction for attribution-focused research.

In addition to this decision process, we also observed several challenges for APT attribution in practice, which require further investigation. These challenges broadly are caused by existing tools being unable to support more diverse and complex data, as well as issues in collaboration, internally and externally, which is essential for the more complex levels of attribution. In this section, we provide recommendations for attribution tool development and threat intelligence sharing based on our results and discuss how existing efforts could be improved through these recommendations.

8.1 Recommendations for Tool Development

Identifying *what* should have priority over identifying *who*. We identified that practitioners are often less concerned with the specific entity behind an attack and more focused on determining that entity's typical TTPs and motivations. By understanding what the adversary might do, the organization can focus their incident response on systems and indicators associated with those TTPs and prioritize threats that pose the greatest risk to their organization.

Practitioners in the field express a need for more automated techniques that can accurately cluster TTPs to provide actionable insights. As P8IR/MA explained it was most important to have automation that gave them a starting point saying they wanted automation that would "give me a hint to work with about the attacker and then I can manually do some work... nowadays you don't have any where to start." Automating TTP-level attribution helps analysts quickly identify the relevant tactics and guide their response efforts. It serves as an initial filter, allowing analysts to narrow down the possibilities of TTPs and associated threat actors. Future research should prioritize identifying *what* threat actors are likely to do rather than focusing on *who* they are. This approach involves mapping low-level threat events to TTPs and correlating them with clusters of known TTPs using frameworks like MITRE ATT&CK. Moreover, TTP-based attribution should be interpretable, providing a high-level summary of the attack and guiding analysts in understanding the scope and magnitude of the incident. By shifting the focus towards TTP-level attribution and ensuring that these systems are computationally feasible and applicable in real-world scenarios, we can better equip practitioners to handle complex threats.

Existing attribution approaches primarily aim to identify the APT groups responsible for an incident [25, 56, 58, 69]. Focusing solely on group-based attribution can lead to missing TTPs, especially if a group modifies its tactics. Instead by emphasizing commonalities across TTPs, even when tactics vary slightly between incidents, we can develop a more robust understanding of the specific methods used. This approach can identify nuances in how particular tactics are executed, which could otherwise be overlooked when analyzing across different adversaries.

Some research has advanced in mapping low-level events to TTPs for APT detection, with systems like HOLMES [48] and APTHunter [40] employing provenance analysis and mapping alerts to TTPs using the MITRE ATT&CK framework [51]. However, these approaches have limitations: (1) They rely on cumulative threat information from all attack stages, assuming that an APT attack completes the entire chain, and (2) they are evaluated with synthetic datasets in laboratory settings. These limitations hinder the adoption of these systems in real-world settings. Our study indicates that practitioners often lack a complete view of the attack chain initially and only uncover the full APT attack sequence through in-depth forensic analysis (see Section 6.2). Future research should build on existing APT attribution approaches by incorporating TTP coverage as a core metric. The effectiveness of such systems should be measured by their accuracy in identifying the correct TTPs, enabling analysts to get a comprehensive understanding of APT-level incidents.

Malware-based APT attribution demands a shift from basic binary clustering to include diverse artifacts. Attributing APTs is inherently complex, necessitating a multi-layered approach and the investigation of extensive data. As APTs increase in sophistication, reliance on infrastructure features becomes less reliable (see Section 6.4). The use of legitimate tools by APTs further complicates detection efforts, and the lack of tools for analyzing diverse file formats increases manual effort (see Section 7.1), a challenge also highlighted in our study on analyzing malicious documents [61].

Practitioners emphasize that analyzing individual artifacts in “gray areas” and understanding their connections to other components in an APT attack chain is important for assessing maliciousness. Malware clustering and classification solutions [1, 9, 27, 46, 50, 53, 57] have been researched for decades. However, current malware-based attribution practices, which focus primarily on classifying binary samples, often fail to provide comprehensive insights into APT-level activities. To enhance attribution depth, it is essential to incorporate a broader range of suspicious artifacts from the APT attack chain, such as phishing lures used to deploy malware and the exploitation of native binaries, such as PowerShell, and associated scripts. Our recent work ADAPT [60] demonstrates progress in this area by incorporating diverse file attributes for APT campaign and group attribution, highlighting the need to account for the heterogeneous artifacts in APT attack chains. Further, ADAPT employs features from secondary sources such as YARA rules and attributes from internet scanning databases such as Censys to augment malware-related features. Future research should build on these studies and develop robust automation for analyzing and extracting indicators from a diverse range of file types.

8.2 Recommendations for Threat Intelligence

Addressing the lack of standardization requires a combination of automated tools, manual review processes, and community collaboration. One of the major challenges in APT attribution is the inconsistent naming and labeling of threat actors and their associated TTPs [22]. As highlighted by our participants in Section 7.2, different organizations, research groups, and governments may assign different names or labels to the same APT group or activity based on their independent analyses, complicating the process of merging and correlating threat data, leading to delays in response efforts.

To address this challenge, future work should explore the development of a comprehensive registry that standardizes the naming conventions for APT groups and their associated TTPs. This registry could leverage existing automated relabeling approaches aimed at reclassifying and standardizing labeling for malware families [62]. To complement this automation, the registry could integrate analyst feedback to enhance potential mappings by using clustering techniques that suggest possible mappings between different naming conventions. These systems would not only propose mappings but also explain the similarities between different labels, helping analysts understand the explanation behind the proposed unification. A key feature of this registry would be its ability to facilitate manual review and validation. By allowing threat analysts to compare samples and naming conventions, a common challenge noted by our participants (see Section 7.2), the tool would help resolve attribution discrepancies and ensure that the standardization aligns with community consensus.

The issue of naming standardization is further complicated by the potential involvement of government agencies, which may have access to unique intelligence resources. As P4IR pointed out “Maybe these governments are doing their own attribution. Maybe they’re like, oh, Mandiant, FireEye, Citizen Lab, like whatever. We’re not going to even consider that. We’re just going to go to our friends at the NSA who have this global view of the internet and ask them to do the attribution. Right. Like, yeah. So, I don’t know, it it’s always difficult to understand whether, you know, attribution is sort of being replicated behind the scenes or whether governments, for instance, are using the attribution of, of, you know, threat intelligence companies or groups.” This emphasizes the need for greater transparency and collaboration between public and private sector entities in the attribution process.

Future research on threat intelligence should consider evaluating TTP coverage and accuracy, beyond traditional IoCs to address the complex nature of APT activities. Bowuman et al. [13] and Li et al. [37] explored the effectiveness of paid threat intelligence and OSINT by looking at the coverage, accuracy, and timeliness of the information presented in them. Specifically, they looked at the IoCs such as domain names, IP addresses, and file hashes. However, in our study, participants reported the unreliability of these features because of a lack of specificity and susceptibility to evasion techniques employed by sophisticated threat actors (see Section 7.1). The reliance on these weak IoCs does not adequately address the complexity of APT activities, which often involve sophisticated TTPs beyond simple indicators.

While some progress has been made, particularly in automating the extraction of IoCs from unstructured text [38] and TTPs from CTI reports [4] future research should build on these studies by emphasizing the evaluation of TTP coverage and accuracy associated with threat actors. This involves assessing how well current threat intelligence solutions capture and represent the complex behaviors and capabilities of APTs. Additionally, further research should focus on developing metrics to measure TTP coverage across different threat intelligence sources.

Acknowledgments

This material is based on research supported by the Vienna Science and Technology Fund (WWTF) and the City of Vienna [10.47379/ICT19056], the SecInt Doctoral College at TU Wien, and SBA Research (SBA-K1), a COMET Center within the COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMAW, and the federal state of Vienna. The COMET Programme is managed by FFG. This work has also been partially supported by NSF grant CNS-2247954, the MATM project (C127/23), with the collaboration of the Spanish Instituce for Cybesecurity (INCIBE), and the Recovery, Transformation and Resilience Plan funded by the European Union (Next Generation).

Ethics considerations

This study was performed in collaboration between institutions in Europe and the US and was approved by the TU Wien's Research Ethics Committee (REC) in Europe and the Institutional Review Board (IRB) at Tufts University in the US. Informed consent was obtained from all participants during the screening survey, and they were provided with detailed information about the research objectives and interview protocol. While we collected email addresses for interview scheduling purposes, this was the only personally identifiable information (PII) gathered. These email addresses were deleted once no longer needed, and they were not stored with the interview data. During transcription, all names of individuals, countries, and organizations mentioned were anonymized using unique identifiers to protect participant identities. The consent form clearly stated the use of transcription service, ensuring compliance with GDPR rules involving third-party access to recorded data [45]. Participants were given the option to conduct interviews without video if it made them more comfortable, and they could choose not to show their faces during video calls. Additionally, if participants shared their screens to demonstrate workflows, we ensured this information was kept secure and not shared further. The data analysis was conducted on the first author's institution premises, and only aggregated results and anonymized transcripts were shared among the research team. To prevent the potential misuse of research data and address risks associated with disclosing information about threat actors, we encouraged participants to withhold or omit sensitive details they were uncomfortable sharing, particularly if they felt such information could be exploited by malicious parties. Participants were also allowed to speak "off-the-record" by pausing the recording at any time. Participants were given the opportunity to review any quotes attributed to them before publication and the context for specific quotes was provided by describing the respondent's role and sector.

One potential concern with publishing this work is that malicious actors could exploit our findings to target organizations by leveraging known gaps in APT response and attribution strategies. However, we consider this risk to be minimal, as the security community has a general idea of these challenges of detecting sophisticated threats [2, 22]. Additionally, detailed threat reports on the operations of various APT groups have already been published by researchers [17, 35, 65]. We believe our work makes a significant contribution to the research community by enhancing the understanding of APT classification and attribution efforts. To the best of our knowledge, current research lacks a comprehensive understanding from security practitioners on how they investigate and attribute APT attacks and the unique challenges they face. Beyond understanding our participants' processes and practices we further explore the challenges they encounter. This helps us provide actionable recommendations for future research efforts in improving and enhancing attribution methods and tools. Therefore, we believe that the benefits of our research—specifically, the advancement of knowledge and practice in analyzing APT-level incidents—outweigh the potential risks.

Open Science

To support transparency, replication, future research, and compliance with the open science policy, we include relevant research materials as part of our artifact, namely (1) interview questions, (2) survey questions, and (3) a codebook, at https://osf.io/hjdk2/.

To comply with data protection requirements and maintain ethical research practices, we do not include raw interview data, such as audio recordings or transcripts, in our replication package. This decision reflects our strong commitment to safeguarding participant privacy and ensuring their right to data protection. By excluding this data, we mitigate the risk of inadvertently disclosing information that could potentially identify our participants or their roles. Instead, we present our findings using thematic analysis and anonymized interview quotes, ensuring our research insights are shared without compromising participant confidentiality.

References

[1] Mansour Ahmadi, Dmitry Ulyanov, Stanislav Semenov, Mikhail Trofimov, and Giorgio Giacinto. "Novel Feature Extraction, Selection and Fusion for Effective Malware Family Classification". In: *Proceedings of the 6th ACM Conference on Data and Application Security and Privacy (CODASPY)*. 2016. DOI: https://doi.org/10.1145/2857705.2857713.

[2] Olusola Akinrolabu, Ioannis Agrafiotis, and Arnau Erola. "The Challenge of Detecting Sophisticated Attacks: Insights from SOC Analysts". In: *Proceedings of the 13th International Conference on Availability, Reliability and Security (ARES)*. 2018. DOI: https://doi.org/10.1145/3230833.3233280.

[3] Bushra A. Alahmadi, Louise Axon, and Ivan Martinovic. "99% False Positives: A Qualitative Study of SOC Analysts' Perspectives on Security Alarms". In: *Proceedings of the 31st USENIX Security Symposium (USENIX Sec)*. 2022.

[4] Md. Tanvirul Alam, Dipkamal Bhusal, Youngja Park, and Nidhi Rastogi. "Looking beyond IoCs: Automatically Extracting Attack Patterns from External CTI". In: *Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. 2023. DOI: https://doi.org/10.1145/3607199.3607208.

[5] Md. Monowar Anjum, Shahrear Iqbal, and Benoit Hamelin. “ANUBIS: A Provenance Graph-Based Framework for Advanced Persistent Threat Detection”. In: *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing (SAC)*. 2022. DOI: https://doi.org/10.1145/3477314.3507097.

[6] John Annett. “Hierarchical Task Analysis”. In: *Handbook of Cognitive Task Design*. 2003.

[7] Simone Aonzo, Yufei Han, Alessandro Mantovani, and Davide Balzarotti. “Humans vs. Machines in Malware Classification”. In: *Proceedings of the 32nd USENIX Security Symposium (USENIX Sec)*. 2023.

[8] Priyanka Badva, Kopo M. Ramokapane, Eleonora Pantano, and Awais Rashid. “Unveiling the Hunter-Gatherers: Exploring Threat Hunting Practices and Challenges in Cyber Defense”. In: *Proceedings of the 33rd USENIX Security Symposium (USENIX Sec)*. 2024.

[9] Ulrich Bayer, Paolo Milani Comparetti, Clemens Hlauschek, Christopher Kruegel, and Engin Kirda. “Scalable, Behavior-based Malware Clustering”. In: *Proceedings of the 16th Network and Distributed System Security Symposium (NDSS)*. 2009.

[10] David Bianco. *The Pyramid of Pain*. https://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html. 2014.

[11] *Binary Ninja*. https://binary.ninja/. 2024.

[12] Marcus Botacin. “What do Malware Analysts want from Academia? A Survey on the State-of-the-practice to Guide Research Developments”. In: *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. 2024. DOI: https://doi.org/10.1145/3678890.3678892.

[13] Xander Bouwman, Harm Griffioen, Jelle Egbers, Christian Doerr, Bram Klievink, and Michel Van Eeten. “A Different Cup of TI? The Added Value of Commercial Threat Intelligence”. In: *Proceedings of the 29th USENIX Security Symposium (USENIX Sec)*. 2020.

[14] Rebekah Brown and Pasquale Stirparo. *SANS 2022 Cyber Threat Intelligence Survey*. 2022. URL: https://www.sans.org/white-papers/sans-2022-cyber-threat-intelligence-survey/.

[15] Rufus Brown, Van Ta, Douglas Bienstock, Geoff Ackerman, and John Wolfram. *Does This Look Infected? A Summary of APT41 Targeting U.S. State Governments*. https://www.mandiant.com/resources/blog/apt41-us-state-governments. 2022.

[16] Juliet Corbin and Anselm Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, 2014. DOI: https://doi.org/10.4135/9781452230153.

[17] Dan Mcwhorter. *APT1: Exposing One of China’s Cyber Espionage Units*. https://www.mandiant.com/resources/apt1-exposing-one-of-chinas-cyber-espionage-units. 2021.

[18] Constanze Dietrich, Katharina Krombholz, Kevin Borgolte, and Tobias Fiebig. “Investigating System Operators’ Perspective on Security Misconfigurations”. In: *Proceedings of the 25th ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 2018. DOI: https://doi.org/10.1145/3243734.3243794.

[19] Recorded Future. *Triage*. https://tria.ge/. 2024.

[20] Ibrahim Ghafir, Mohammad Hammoudeh, Vaclav Prenosil, Liangxiu Han, Robert Hegarty, Khaled Rabie, and Francisco J Aparicio-Navarro. “Detection of Advanced Persistent Threat using Machine-Learning Correlation Analysis”. In: *Future Generation Computer Systems* 89 (2018). DOI: https://doi.org/10.1016/j.future.2018.06.055.

[21] John R. Goodall, Wayne G. Lutters, and Anita Komlodi. “I Know my Network: Collaboration and Expertise in Intrusion Detection”. In: *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work (CSCW)*. 2004. DOI: https://doi.org/10.1145/1031607.1031663.

[22] Jason Gray, Daniele Sgandurra, Lorenzo Cavallaro, and Jorge Blasco. “Identifying Authorship in Malicious Binaries: Features, Challenges & Datasets”. In: *ACM Computing Surveys (CSUR)* 56.8 (2024). DOI: https://doi.org/10.1145/3653973.

[23] Harm Griffioen, Tim Booij, and Christian Doerr. “Quality Evaluation of Cyber Threat Intelligence Feeds”. In: *Proceedings of the 18th International Conference of Applied Cryptography and Network Security (ACNS)*. 2020. DOI: https://doi.org/10.1007/978-3-030-57878-7_14.

[24] Greg Guest, Arwen Bunce, and Laura Johnson. “How Many Interviews are Enough? An Experiment with Data Saturation and Variability”. In: *Field Methods* 18.1 (2006). DOI: 10.1177/1525822X05279903.

[25] Weijie Han, Jingfeng Xue, Yong Wang, Fuquan Zhang, and Xianwei Gao. “APTMalInsight: Identify and Cognize APT Malware Based on System Call Information and Ontology Knowledge Framework”. In: *Information Sciences* 546 (2021). DOI: https://doi.org/10.1016/j.ins.2020.08.095.

[26] Xueyuan Han, Thomas Pasquier, Adam Bates, James Mickens, and Margo Seltzer. “UNICORN: Runtime Provenance-Based Detector for Advanced Persistent Threats”. In: *Proceedings of the 27th Network and Distributed System Security Symposium (NDSS)*. 2020.

DOI: https://doi.org/10.14722/ndss.2020.24046.

[27] Mehadi Hassen, Marco M. Carvalho, and Philip K. Chan. “Malware Classification using Static Analysis-based Features”. In: *Proceedings of the IEEE Symposium Series on Computational Intelligence (SSCI)*. 2017. DOI: https://doi.org/10.1109/ssci.2017.8285426.

[28] Andrew F Hayes and Klaus Krippendorff. “Answering the Call for a Standard Reliability Measure for Coding Data”. In: *Communication Methods and Measures* 1.1 (2007). DOI: https://doi.org/10.1080/19312450709336664.

[29] *IDA Pro*. https://hex-rays.com/ida-pro/. 2024.

[30] Zian Jia, Yun Xiong, Yuhong Nan, Yao Zhang, Jinjing Zhao, and Mi Wen. “MAGIC: Detecting Advanced Persistent Threats via Masked Graph Representation Learning”. In: *Proceedings of the 32nd USENIX Security Symposium (USENIX Sec)*. 2023.

[31] *JoeSandbox*. https://joesandbox.com/. 2024.

[32] Faris Bugra Kokulu, Ananta Soneji, Tiffany Bao, Yan Shoshitaishvili, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn. “Matched and Mismatched SOCs: A Qualitative Study on Security Operations Center Issues”. In: *Proceedings of the 26th ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 2019. DOI: https://doi.org/10.1145/3319535.3354239.

[33] Jesse Kornblum. “Identifying Almost Identical Files using Context Triggered Piecewise Hashing”. In: *Digital Investigation* 3 (2006). DOI: https://doi.org/10.1016/j.diin.2006.06.015.

[34] Marc Kührer, Christian Rossow, and Thorsten Holz. “Paint it Black: Evaluating the Effectiveness of Malware Blacklists”. In: *Proceedings of the 17th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. 2014. DOI: https://doi.org/10.1007/978-3-319-11379-1_1.

[35] Clement Lecigne and Maddie Stone. *Active North Korean Campaign Targeting Security Researchers*. https://blog.google/threat-analysis-group/active-north-korean-campaign-targeting-security-researchers/. 2023.

[36] Shaofei Li, Feng Dong, Xusheng Xiao, Haoyu Wang, Fei Shao, Jiedong Chen, Yao Guo, Xiangqun Chen, and Ding Li. “NODLINK: An Online System for Fine-Grained APT Attack Detection and Investigation”. In: *Proceedings of the 31st Network and Distributed System Security Symposium (NDSS)*. 2024. DOI: https://doi.org/10.14722/ndss.2024.23204.

[37] Vector Guo Li, Matthew Dunn, Paul Pearce, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage. “Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence”. In: *Proceedings of the 28th USENIX Security Symposium (USENIX Sec)*. 2019.

[38] Xiaojing Liao, Kan Yuan, XiaoFeng Wang, Zhou Li, Luyi Xing, and Raheem Beyah. “Acing the IoC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence”. In: *Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 2016. DOI: https://doi.org/10.1145/2976749.2978315.

[39] Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. “Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability”. In: *Human Communication Research* 28.4 (2002). DOI: https://doi.org/10.1093/hcr/28.4.587.

[40] Moustafa Mahmoud, Mohammad Mannan, and Amr Youssef. “APTHunter: Detecting Advanced Persistent Threats in Early Stages”. In: *Digital Threats: Research and Practice* 4 (2022). DOI: https://doi.org/10.1145/3559768.

[41] Alessandro Mantovani, Simone Aonzo, Yanick Fratantonio, and Davide Balzarotti. “RE-Mind: A First Look inside the Mind of a Reverse Engineer”. In: *Proceedings of the 31st USENIX Security Symposium (USENIX Sec)*. 2022.

[42] Morgan Marquis-Boire, Marion Marschalek, and Claudio Guarnieri. “Big Game Hunting: The Peculiarities in Nation-State Malware Research”. In: *BlackHat USA*. 2015.

[43] James Mattei, Madeline McLaughlin, Samantha Katcher, and Daniel Votipka. “A Qualitative Evaluation of Reverse Engineering Tool Usability”. In: *Proceedings of the 38th Annual Computer Security Applications Conference (ACSAC)*. 2022. DOI: https://doi.org/10.1145/3564625.3567993.

[44] William P Maxam III and James C Davis. “An Interview Study on Third-Party Cyber Threat Hunting Processes in the US Department of Homeland Security”. In: *Proceedings of the 33rd USENIX Security Symposium (USENIX Sec)*. 2024.

[45] *MAXQDA*. https://www.maxqda.com/automatic-transcription. 2024.

[46] Francesco Meloni, Alessandro Sanna, Davide Maiorca, and Giorgio Giacinto. “Effective Call Graph Fingerprinting for the Analysis and Classification of Windows Malware”. In: *Proceedings of 19th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*. 2022. DOI: https://doi.org/10.1007/978-3-031-09484-2_3.

[47] Leigh Metcalf and Jonathan M. Spring. “Blacklist Ecosystem Analysis”. In: *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*. 2015. DOI: https://doi.org/10.1145/2808128.2808129.

[48] Sadegh M Milajerdi, Rigel Gjomemo, Birhanu Eshete, Ramachandran Sekar, and V.N. Venkatakrishnan. “Holmes: Real-Time APT Detection through Correlation of Suspicious Information Flows”. In: *Proceedings of the 40th IEEE Symposium on Security & Privacy (S&P)*. 2019. DOI: https://doi.org/10.1109/sp.2019.00026.

[49] Jaron Mink, Hadjer Benkraouda, Limin Yang, Arridhana Ciptadi, Ali Ahmadzadeh, Daniel Votipka, and Gang Wang. “Everybody’s Got ML, Tell Me What Else You Have: Practitioners’ Perception of ML-Based Security Tools and Explanations”. In: *Proceedings of the 44th IEEE Symposium on Security & Privacy (S&P)*. 2023. DOI: https://doi.org/10.1109/sp46215.2023.10179321.

[50] Omid Mirzaei, Roman Vasilenko, Engin Kirda, Long Lu, and Amin Kharraz. “Scrutinizer: Detecting Code Reuse in Malware via Decompilation and Machine Learning”. In: *Proceedings of the 18th Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*. 2021. DOI: https://doi.org/10.1007/978-3-030-80825-9_7.

[51] MITRE Corporation. *ATT&CK Matrix*. https://attack.mitre.org/. 2024.

[52] Sean Oesch, Robert Bridges, Jared Smith, Justin Beaver, John Goodall, Kelly Huffer, Craig Miles, and Dan Scofield. “An Assessment of the Usability of Machine Learning Based Tools for the Security Operations Center”. In: *Proceedings of the International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. 2020. DOI: https://doi.org/10.1109/ithings-greencom-cpscom-smartdata-cybermatics50389.2020.00111.

[53] Avi Pfeffer, Catherine Call, John Chamberlain, Lee Kellogg, Jacob Ouellette, Terry Patten, Greg Zacharias, Arun Lakhotia, Suresh Golconda, John Bay, Robert Hall, and Daniel Scofield. “Malware Analysis and Attribution using Genetic Information”. In: *Proceedings of the 7th International Conference on Malicious and Unwanted Software (MALWARE)*. 2012. DOI: https://doi.org/10.1109/malware.2012.6461006.

[54] Emilee Rader, Samantha Hautea, and Anjali Munasinghe. ““I Have a Narrow Thought Process”: Constraints on Explanations Connecting Inferences and Self-Perceptions”. In: *Proceedings of the 16th Symposium On Usable Privacy and Security (SOUPS)*. 2020.

[55] Chris Ray. *Intro to ImpHash for DFIR: "Fuzzy" Malware Matching*. https://www.cybertriage.com/blog/intro-to-imphash-for-dfir-fuzzy-malware-matching/. 2024.

[56] Yitong Ren, Yanjun Xiao, Yinghai Zhou, Zhiyong Zhang, and Zhihong Tian. “CSKG4APT: A Cybersecurity Knowledge Graph for Advanced Persistent Threat Organization Attribution”. In: *IEEE Transactions on Knowledge and Data Engineering* 35 (2022). DOI: https://doi.org/10.1109/tkde.2022.3175719.

[57] Konrad Rieck, Thorsten Holz, Carsten Willems, Patrick Düssel, and Pavel Laskov. “Learning and Classification of Malware Behavior”. In: *Proceedings of the 5th International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*. 2008. DOI: https://doi.org/10.1007/978-3-540-70542-0_6.

[58] Ishai Rosenberg, Guillaume Sicard, and Eli Omid David. “DeepAPT: Nation-State APT Attribution Using End-to-End Deep Neural Networks”. In: *Proceedings of the 26th International Conference on Artificial Neural Networks (ICANN)*. 2017. DOI: https://doi.org/10.1007/978-3-319-68612-7_11.

[59] Vinay Sachidananda, Rajendra Patil, Akshay Sachdeva, Kwok-Yan Lam, and Liu Yang. “APTer: Towards the Investigation of APT Attribution”. In: *Proceedings of the 6th IEEE Conference on Dependable and Secure Computing (DSC)*. 2023. DOI: https://doi.org/10.1109/dsc61021.2023.10354155.

[60] Aakanksha Saha, Jorge Blasco, Lorenzo Cavallaro, and Martina Lindorfer. “ADAPT it! Automating APT Campaign and Group Attribution by Leveraging and Linking Heterogeneous Files”. In: *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*. 2024. DOI: https://doi.org/10.1145/3678890.3678909.

[61] Aakanksha Saha, Jorge Blasco, and Martina Lindorfer. “Exploring the Malicious Document Threat Landscape: Towards a Systematic Approach to Detection and Analysis”. In: *Proceedings of the 3rd Workshop on Rethinking Malware Analysis (WoRMA)*. 2024. DOI: https://doi.org/10.1109/eurospw61312.2024.00065.

[62] Silvia Sebastián and Juan Caballero. “AVclass2: Massive Malware Tag Extraction from AV Labels”. In: *Proceedings of the 36th Annual Computer Security Applications Conference (ACSAC)*. 2020. DOI: https://doi.org/10.1145/3427228.3427261.

[63] Sathya Chandran Sundaramurthy, Alexandru G Bardas, Jacob Case, Xinming Ou, Michael Wesch, John McHugh, and S Raj Rajagopalan. “A Human Capital Model for Mitigating Security Analyst Burnout”. In: *Proceedings of the 11th Symposium On Usable Privacy and Security (SOUPS)*. 2015.

[64] Wiem Tounsi and Helmi Rais. “A Survey on Technical Threat Intelligence in the Age of Sophisticated Cyber Attacks”. In: *Computers & Security* 72 (2018). DOI: https://doi.org/10.1016/j.cose.2017.09.001.

[65] Unit42. *Russia’s Gamaredon aka Primitive Bear APT Group Actively Targeting Ukraine*. https://unit42.paloaltonetworks.com/gamaredon-primitive-bear-ukraine-update-2021/. 2022.

[66] *VirusTotal*. https://www.virustotal.com/. 2024.

[67] Daniel Votipka, Seth Rabin, Kristopher Micinski, Jeffrey S. Foster, and Michelle L. Mazurek. “An Observational Investigation of Reverse Engineers’ Processes”. In: *Proceedings of the 29th USENIX Security Symposium (USENIX Sec)*. 2020.

[68] Daniel Votipka, Rock Stevens, Elissa Redmiles, Jeremy Hu, and Michelle Mazurek. “Hackers vs. Testers: A Comparison of Software Vulnerability Discovery Processes”. In: *Proceedings of the 39th IEEE Symposium on Security & Privacy (S&P)*. 2018. DOI: https://doi.org/10.1109/sp.2018.00003.

[69] Qinqin Wang, Hanbing Yan, and Zhihui Han. “Explainable APT Attribution for Malware using NLP Techniques”. In: *Proceedings of the 21st IEEE International Conference on Software Quality, Reliability and Security (QRS)*. 2021. DOI: https://doi.org/10.1109/qrs54544.2021.00018.

[70] Chad Warner. *Diamond Model in Cyber Threat Intelligence*. https://warnerchad.medium.com/diamond-model-for-cti-5aba5ba5585. 2021.

[71] Georg Wicherski. “peHash: A Novel Approach to Fast Malware Clustering”. In: *Proceedings of the 2nd USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*. 2009.

[72] Miuyin Yong Wong, Matthew Landen, Frank Li, Fabian Monrose, and Mustaque Ahamad. “Comparing Malware Evasion Theory with Practice: Results from Interviews with Expert Analysts”. In: *Proceedings of the 20th Symposium On Usable Privacy and Security (SOUPS)*. 2024.

[73] Khaled Yakdan, Sergej Dechand, Elmar Gerhards-Padilla, and Matthew Smith. “Helping Johnny to Analyze Malware: A Usability-Optimized Decompiler and Malware Analysis User Study”. In: *Proceedings of the 37th IEEE Symposium on Security & Privacy (S&P)*. 2016. DOI: https://doi.org/10.1109/sp.2016.18.

[74] Miuyin Yong Wong, Matthew Landen, Manos Antonakakis, Douglas M Blough, Elissa M Redmiles, and Mustaque Ahamad. “An Inside Look into the Practice of Malware Analysis”. In: *Proceedings of the 28th ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 2021. DOI: https://doi.org/10.1145/3460120.3484759.

A Survey Questionnaire

Consent and GDPR Consent. In this section, we obtain informed consent from participants, ensuring they understand the study’s purpose, procedures, data protection measures, and their rights.

Participant Information. In this section, we ask a few questions about your background, name, age, and email address.

Job Description.

1. What is your current job role and job title?
2. Please specify your highest level of education. Less than high school, High School graduate (high school diploma or equivalent such as GED), Some college, but no degree, Associate Degree, Bachelor’s Degree, Master’s Degree, Doctorate Degree (MD, PhD, JD, etc.), Prefer not to answer.
3. Please rate your expertise on the scale of Novice (basic knowledge), Fundamental Awareness (limited experience), Intermediate (practical application), Advanced (applied theory), Expert (recognized authority), None in the area of Malware analysis, APT tracking and attribution, Threat intelligence research, and Incidence response.
4. What is the end goal of your APT threat analysis work? (Please check all that apply) Forensics, Attribution, Classification and clustering, Signature creation, Indicators of Compromise, Research, Writing threat reports, and others.

Work Experience. Please list any tools you use when performing APT malware analysis, APT incidence response, or threat intelligence research. Please continue listing tools as you have a new line for each tool and continue to list tools until you cannot think of any more. These can be any tools you have used, you do not need to regularly use them.

B Interview Questions

Background and Experience.

- Could you tell me a little bit about your job role and experience?
- How did you get into the field of investigating APTs? Could you describe any interesting or recent APT incident that you worked on? What steps did you take to investigate the suspected APT attack?
- What is the end goal of your work pipeline—collecting IoCs, malware analysis, threat intelligence research, or writing threat reports?

APT Analysis Objectives.

- What is the end goal of analyzing an APT sample?
  - Is it to identify tactics, observe behavior, attribute the attack, or something else?
- Where do you source your APT samples from?
  - How do you ensure that the samples are relevant to your study?
  - Do you use in-house SIEM systems, collect samples from VT or through clients/repositories, or use other methods?
- When do you start considering a malicious sample as part of a suspected APT campaign or operation?
  - How do you prioritize samples for further analysis based on their potential association with known APT campaigns?
  - Are there specific indicators or techniques you use for this prioritization?

Process and Pipeline.

- What is your digital forensics and incident response protocol when you identify malicious activity as part of an APT campaign?
  - Do you have separate workflows for dealing with APTs versus traditional malware threats?
- What is your process for attributing a sample to a specific threat group?
  - What features do you consider when making this attribution?
- For a newly identified APT campaign, what is your process for correlating samples with previously established campaigns?
  - How do you identify similar patterns or connections between APT campaigns?
  - Do you use ML-based automation for sample correlation?
  - If yes, do you see any challenges in working with ML-based tools?
- What is your approach to gathering comprehensive information about APT groups?
  - How do you keep track of potentially related campaigns over time, especially when they are spread across multiple sources?
- Are there challenges in incorporating threat intelligence into your analysis of APT campaigns?
  - What challenges do you face when attributing attacks to specific groups?
  - Are there any other challenges when grouping attacks that may have been carried out by APT groups with aliases or changing names?
- How do you effectively aggregate and consolidate data from diverse OSINT sources?
  - Can you provide an example of this process?
- What techniques do you use to manage publicly available information about APT campaigns?
  - How do you detect and eliminate redundancies?

Final Remarks.

- What are the primary concerns when dealing with APT incidents?
  - Are accuracy and precision prioritized, or is there more focus on speed, automation, or developing a generic framework for file types?
- What processes, tools, and policies currently work best in your team?