An Interview Study on Third-Party Cyber Threat Hunting Processes in the U.S. Department of Homeland Security

William P. Maxam III
United States Coast Guard Academy[*]

James C. Davis
Purdue University

Abstract

Cybersecurity is a major challenge for large organizations. Traditional cybersecurity defense is reactive. Cybersecurity operations centers keep out adversaries and incident response teams clean up after break-ins. Recently a proactive stage has been introduced: Cyber Threat Hunting (TH) looks for potential compromises missed by other cyber defenses. TH is mandated for federal executive agencies and government contractors. As threat hunting is a new cybersecurity discipline, most TH teams operate without a defined process. The practices and challenges of TH have not yet been documented.

To address this gap, this paper describes the first interview study of threat hunt practitioners. We obtained access and interviewed 11 threat hunters associated with the U.S. government's Department of Homeland Security. Hour-long interviews were conducted. We analyzed the transcripts with process and thematic coding. We describe the diversity among their processes, show that their processes differ from the TH processes reported in the literature, and unify our subjects' descriptions into a single TH process. We enumerate common TH challenges and solutions according to the subjects. The two most common challenges were difficulty in assessing a Threat Hunter's expertise, and developing and maintaining automation. We conclude with recommendations for TH teams (improve planning, focus on automation, and apprentice new members) and highlight directions for future work (finding a TH process that balances flexibility and formalism, and identifying assessments for TH team performance).

1 Introduction

Computer network security is a challenge in the modern world. Cyber intrusions are a concern for both governments and private corporations. Unauthorized network infiltrations cost individual organizations an average of $13 million a year [8] and may compromise their operations or intellectual property. Governments have additional non-monetary concerns, such as protecting election systems and maintaining national security [7]. The longer an adversary dwells undetected on a network, the more damage the adversary can cause. One analysis found that a 50% reduction in dwell time would reduce the cost of an attack by ∼30% [36]. IBM found that data breaches cost on average $1.12 million more if not contained within the first 200 days, with costs including lost revenue, regulatory and legal fees, and forensics activities [9]. They estimate the average adversary-dwell-time at 230 days, not including additional time to respond to the breach [9].

The primary method of finding undetected network intruders is a Cyber Threat Hunt (TH). Threat Hunting is "*a focused and iterative approach to searching out, identifying, and understanding adversaries internal to the defender's networks*" [60]. The government routinely performs **third-party hunting**, *i.e.,* TH on contractors' networks [7]. The private sector conducts threat hunting as well [24], more commonly within their own organizations. According to a 2017 SANS Institute survey, across sectors including telecommunications, technology, government, healthcare, and finance, most organizations engage in threat hunting but in an immature way [59]. Under half of the organizations had a defined TH process [59].

The processes currently used by TH teams are not well documented [94]. Prior research on TH teams does not describe TH processes in detail [25,37,92,95] or is focused on internal hunt teams [94]. This knowledge gap limits a team's ability to adopt best practices and improve their processes over time, and it also limits how well researchers can assist TH teams.

To address the lack of TH process understanding, we interviewed professional threat hunters. We used a semi-structured interview methodology, selected as a result of the data available from previous work done in the TH domain and an anticipated small sample size (§3.3). We interviewed 11 TH practitioners from two organizations within the US Department of Homeland Security (DHS). Each subject was interviewed for ∼1 hour. We combined subjects' TH process sketches and dialogue to create a unified process model of the US DHS threat hunt process. We also thematically coded the transcripts to elicit subjects' problems and solutions.

[*] Work performed while at Purdue University.

Our results provide researchers with a better understanding of TH processes. We provide the first academic description of the cybersecurity landscape to include TH teams (§2) and the first published process model describing DHS TH teams (Figure 2). Unlike prior literature that recommends a hypothesis-driven process, we report that the studied TH teams incorporate a *data-driven* process. We also document the challenges that practitioners report with this process, and describe the best practices they recommend (Table 3). Their open questions are captured as future work (§5).

Our contributions are:

- We provide an updated description of the cybersecurity landscape that includes Threat Hunting teams (§2).
- We characterize US DHS Threat Hunt processes, both in comparison to prior work and via a novel process induction (§4.1 and Figure 2).
- We describe problems, possible solutions, and open questions discussed by TH practitioners (§4.2 and Table 3).

2 Background and Related Work

In §2.1 we describe the landscape of current cybersecurity defenses and the role of Threat Hunting (TH). §2.2 summarizes what is known about public and private TH teams. §2.3 describes common TH frameworks and processes. As there is little academic literature on Threat Hunt, this section is expository. We rely in part on reputable "grey literature".

2.1 Layers of cybersecurity defense

Figure 1 illustrates typical layers of cybersecurity defense as they interact with an adversary. We discuss each layer in turn.

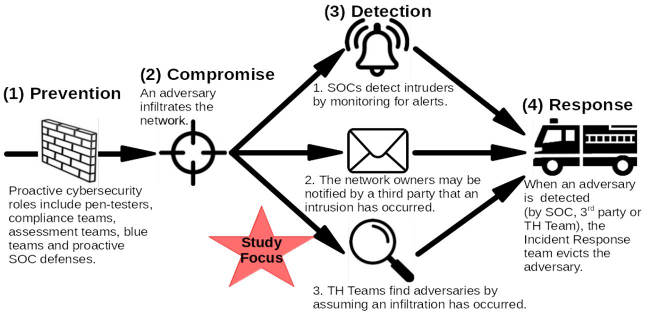
(3) Detection
(2) Compromise
An adversary infiltrates the network.
(1) Prevention
1. SOCs detect intruders by monitoring for alerts.
(4) Response
Proactive cybersecurity roles include pen-testers, compliance teams, assessment teams, blue teams and proactive SOC defenses.
2. The network owners may be notified by a third party that an intrusion has occurred.
When an adversary is detected (by SOC, 3rd party or TH Team), the Incident Response team evicts the adversary.
Study Focus
3. TH Teams find adversaries by assuming an infiltration has occurred.

Figure 1: Threat hunt is one of three common ways to discover an adversary once they have circumvented cyber defenses. Once an adversary is discovered, an IR team responds.

2.1.1 Prevention (Diagram Step 1)

As Figure 1 shows, *preventative* cybersecurity teams are the first line of defense against adversaries. These teams include traditional defensive teams such as Blue Teams, Compliance Teams, and Security Operations Centers (SOCs). Blue teams seek to protect the network from intruders by improving the network's security [98]. Compliance Teams enforce cybersecurity best practices across a network [43]. SOCs are primarily responsible for detecting adversaries as they attempt to gain access to a network [26]. SOCs in particular have received much attention from researchers. Researchers have interviewed SOC members [39] and embedded in SOCs [88]. Classification systems exist for SOCs to measure their maturity and capability [54], best practices have been enumerated [62] and SOC processes are being studied [78].

There are also offensive preventative security teams. They include Red Teams and Penetration Testing Teams [44]. These teams take offensive actions in order to simulate adversaries. Penetration Testing Teams typically look only for vulnerabilities at the network edge, while red teams seek to infiltrate past the network edge and further into the network.

2.1.2 Compromises (Diagram Step 2)

In Step 2 of Figure 1, the adversary evades these countermeasures and *compromises* the network. If an adversary compromises a network without detection, they can cause significant damage [36]. The longer adversaries remain undetected, the more costly the intrusion becomes [9, 36]. Attackers constantly innovate their capabilities, and many works have described compromises such as SolarWinds [27], Stuxnet [63], and the 2015 Ukrainian power grid attack [99].

2.1.3 Detection (Diagram Step 3)

Step 3 of Figure 1 shows three ways in which such an adversary may subsequently be *detected*. The upper route depicts the SOC detecting the adversary, *e.g.,* via one of the SOC's internal network or host-based sensors. The middle route shows a notification by an agency such as the US FBI. The third path is *proactive discovery*, *i.e.,* Threat Hunt, our study's focus.

TH teams perform a unique function, hunting for adversaries internal to the network boundary. To identify undetected adversaries on the network, TH teams search for adversaries that have evaded detection by usual methods [60]. This is analogous to a military unit relying both on gate guards (the SOC) and also patrols inside the camp to "hunt" for adversaries that breach the gate. TH teams do not configure network defences (as blue teams and compliance teams do), nor do they take offensive action (as red teams and penetration testers do) [66]. Threat Hunt teams only search the network for adversaries that existing defenses may have missed [6]. They hunt on the assumption that an adversary has infiltrated the network, when there is no sign of a compromise [37]. ***Unlike the other steps in Figure 1, there is little empirical data about TH team processes, practices, and challenges.***

2.1.4 Response (Diagram Step 4)

The last step in Figure 1 is *response*. Once an adversary is detected, an Incident Response (IR) team is called to respond. Incident Response teams are responsible for evicting adversaries that have been found on the network [72]. IR teams react to adversaries regardless of how they are detected. IR teams have also been studied by academia and government agencies [40]. Researchers have interviewed IR team members [69, 97] and embedded in IR teams [72, 90].

2.2 Threat Hunting Teams and Organizations

In this section we describe how TH is implemented in the private and public sectors. Although TH is a new discipline [76, 94], it is considered an important cybersecurity capability for security practitioners and academia [47, 68]. TH is mandated for Federal Civilian Executive Branch Agencies [15] like the National Aeronautics and Space Administration (NASA) and the Department of the Treasury [21].

Since TH is an emerging discipline of cybersecurity, it is little researched. The government TH process is particularly important to understand and improve because recent US presidential executive orders [15] and the US National Cybersecurity Strategy of 2018 [7] *mandate* the use of (third-party) government hunters on government contractors' networks.

2.2.1 TH in the Private Sector

TH can either be done using personnel internal to the organization or using a third party's TH team. For an internal TH team, organizations either designate SOC personnel to perform hunting or maintain teams dedicated to TH [60]. Some companies offer threat hunting as a service, including Booz Allen Hamilton [20], Crowdstrike [17] and Cisco [19]. Most organizations in the private sector opt for internal TH teams [59], *e.g.,* for privacy or intellectual property protection.

2.2.2 TH in the Public Sector

The US Government uses both civilian and military teams. For example, the Cybersecurity and Infrastructure Security Agency's (CISA) Hunt and Incident Response Team (HIRT) [12] is a team of civilians that performs TH. On the military side, all branches of the US military [91] and some state National Guards [64] have Cyber Protection Teams (CPTs) — some CPTs conduct TH work along with other cybersecurity functions [16]. Government teams are deployed to federal civilian [15], military [2], and even private sector networks [7, 12]. The Department of Homeland Security includes both CISA and the Coast Guard and thus operates both civilian (CISA) and military (Coast Guard) Hunt Teams.

2.2.3 Common Problems Affecting TH Teams

Both government and private sector TH teams have substantial personnel turnover [11, 65, 74]. In cybersecurity organizations, public and private, cybersecurity analysts commonly change jobs every two years [14]. Military organizations often rotate personnel every 2–3 years [34, 70, 82], placing this concern beyond the control of the individual TH teams to address. Good processes have been shown to mitigate the adverse effects of fewer expert personnel [71], helping mitigate the adverse effects of personnel turnover. Our study describes existing processes as a step toward this goal.

2.3 TH Frameworks and Processes

This section presents TH frameworks and processes. A *TH Framework* is a way of organizing information to assist with the task of hunting. A *TH Process* is a way of organizing the tasks associated with Threat Hunting across time. Beyond this material, the technical report (Appendix A) gives illustrations of the frameworks and processes described here.

2.3.1 TH Frameworks

Both academia [37] and private sector [94] documents suggest three popular frameworks: Lockheed Martin's Kill Chain [3], the Mitre ATT&CK framework [28], and the Pyramid of Pain [32]. All three of these frameworks describe adversary activity in a way that assists the defender in categorizing events and focusing their search. The Kill Chain and ATT&CK frameworks outline the steps an attacker takes to carry out a successful attack; ATT&CK goes into greater detail. The Pyramid of Pain instead assesses what information is most valuable for disrupting adversary activity.

Researchers often suppose a framework called the hypothesis method [25, 29, 37, 52, 95]. The hypothesis method is when the TH team outlines a possible intrusion that could have occurred on the network and then tests that hypothesis using the data available. An example hypothesis is: *"Advanced Persistent Threat #0 exploited CVE-2021-44228 to compromise a VPN Server then moved laterally to the domain controller."*

An alternative to hypothesis hunting is *data-driven unstructured hunting* [83, 84, 94]. Here, threat hunters search for adversaries without a hypothesis, guided by statistical or behavioral analysis to identify adversary activity. Some authors consider this as hunting without a process [84]. Others say it is less efficient than the hypothesis method [13, 83].

These frameworks and methods are complementary [94]. For example, a TH team could build a hypothesis using the steps in the Kill Chain framework. The team could then look up the associated techniques using the ATT&CK framework and prioritize the search based on the Pyramid of Pain.

2.3.2 TH Processes

Currently, the most common way to hunt for an adversary is *ad hoc*, *i.e.,* without a formal process [59]. Some TH teams may have privately documented processes, but public descriptions are rare. We describe what is known.

Private sector processes: To address the lack of standardized processes and definitions, four organizations from the Dutch financial sector shared a process known as the Targeted Hunting integrating Threat Intelligence (TaHiTI) process [94].

TaHiTI describes the behavior of internal threat hunters rather than third-party/external hunters, and thus makes assumptions about the network information that the threat hunters have.

The private sector has made other attempts to explain TH processes but none as detailed as TaHiTI. Endgame published *The Endgame Guide to Threat Hunting* [45,81]. They proposed a 4-phase process with 6 hunting steps. Other TH guides did not provide a TH process [10].

These TH process models — TaHiTI and Endgame — are prescriptive, not descriptive. TaHiTI was derived from a private sector round table, describing what should occur, not necessarily what does occur. Endgame is similar.

US government processes: There is no public documentation of the TH process used by Government TH teams. The closest work is by Trent *et al.* [92]. They created a *cognitive* model of activities performed by US Army CPTs. Their goal was not specifically to outline a Threat Hunting *process*, but as CPTs perform TH their findings are somewhat relevant.

Academic perspectives: Previous academic works on TH methodologies do not describe current TH practices. They propose novel approaches to assist TH teams [25,95].

Comparing private sector and govt. TH processes: Comparison between the TH processes proposed by TaHiTI [94] and Trent *et al.* [92] is imperfect for two reasons: (1) TaHiTI was focused on TH while Trent *et al.* covered TH as well as other cybersecurity roles; and and (2) TaHiTI is a process while Trent *et al.* is a cognitive model. Nevertheless, our analysis of these works suggests differences between private sector and government TH teams. TaHiTI is for internal TH teams, while Trent focuses on third-party CPTs. This means that TaHiTI assumes more shared context across missions, *e.g.,* keeping "investigation abstracts" in a backlog for future hunts. In contrast, the external hunts from Trent are more self-contained, with tasks devoted to "Planning and Logistics" and "Closure". Perhaps related, we also observe that TaHiTI is more abstract while the Trent model is more detailed, including tasks such as Forensic analysis as distinct from Host, Malware, and Network analysis. Finally, the TaHiTI TH process uses the hypothesis method, while Trent *et al.* does not indicate a hypothesis (even though generating a hypothesis would be a unique cognitive task).

2.4 Summary and Unknowns

Although threat hunting is a mandated government function, little is known about TH team processes. The most detailed TH process is TaHiTI, a process described by threat hunters from four private sector institutions [94]. However, government TH processes likely differ from private sector ones because government teams often hunt on third-party networks rather than within their own networks [7]. The TH process used by government teams is unknown.

Since most organizations that perform TH do so without a formal process, understanding TH processes outside of the "internal private sector team" context (TaHiTI) will help organizations developing their own TH processes. Typical TH team practices and challenges are also undocumented — learning these would benefit all TH organizations, whether or not they that have a TH process. One specific area of interest is the effect of turnover and integration of newcomers, a concern shared by private and public-sector TH teams.

3 Research Questions and Methodology

Historically, the US government have driven the creation of cybersecurity standards [86]. For example, after the 2015 Office of Personnel Management (OPM) data breach [55], many private organizations sought to improve their processes from shortcomings observed in the OPM process [4,5,57,61]. Describing the government TH process may be beneficial for the same reasons, both as an example and to assist in identifying shortcomings. The government is among the sectors most often targeted by cyberattackers [22]. The US government's cybersecurity defenses, including its TH processes, are thus of great interest. We provide a first view.

3.1 Research Questions

Describing all US government TH processes is beyond the scope of a single study. As a step toward this goal, in this work we examine TH processes from one branch of the government, the US Department of Homeland Security (DHS). In this context, we address two research questions:

- **RQ1:** What processes are used by DHS TH teams?
- **RQ2:** What shortcomings exist with current DHS TH processes and what might be done to alleviate them?

3.2 Statement of Positionality

One of the authors is a former DHS Threat Hunter. This author's background shaped, both directly and indirectly, our study design, recruiting, analysis, and findings [42]. For example, their professional experiences informed our interview protocol, and we recruited their former colleagues as subjects. The researcher's relationship with some of the subjects may have had positive and negative effects: responses could be biased by the relationship, but they may also be enriched because interviewer-subject trust had already been developed. Despite the potential biases, we emphasize that access to TH personnel has been a barrier to research. Capturing aspects of DHS TH experiences is valuable, even if incomplete.

3.3 Study Design

An Interview methodology was chosen for this study as a result of the data available and the population being studied. We felt the prior work of Trent *et al.* [92] and van Os *et al.* [94] provided enough initial structure to frame the collection instrument and initial analysis. Long-form ($\sim$1hr) interviews allowed us to make the most of a relatively small sample size.

Little information exists on the processes used by government Threat Hunting teams, so our study will direct future research.

We considered and decided against other methodologies. It would be difficult to capture the complexity of the TH process through a survey, which would not allow for iteration on the collection instrument to pick up unexpected intricacies [100]. Additionally, since the population of government threat hunters is small and difficult to access, a meaningful sample size for a survey would be difficult to achieve [30]. We also considered a grounded theory methodology, which is suitable when no prior theory or framework exists [58]. In our case, however, we had access to prior work done by private TH teams (TaHiTI) and studies on TH-adjacent cybersecurity teams (Trent et al.) which we used to guide our study.

The TH teams we studied have three roles: *Leadership*, *e.g.*, officers, deal with TH strategic concerns but do not deploy with the team. Team Leads deploy with a hunt team and act as the on-site manager. Analysts deploy and perform the analytic tasks associated with hunting. Each role may have different perspectives on the TH process, challenges, and solutions, so we recruited subjects in each role.

3.4 Recruitment and Subject Demographics

Government Threat Hunting teams are a difficult group of practitioners to study due to the small size of the teams and the sensitive information they often deal with. One author's US government security clearance and previous TH duties allowed us access to these TH team members and helped ensure no sensitive information was disclosed.

Participants were recruited for interviews through the authors' professional network. Emails were sent to TH members with varying experience and roles. At both organizations, subjects were representative of multiple internal teams and analyst populations, including primary organizational divisions. The response rate from direct contacts was 59% (10/17). One additional subject offered to be interviewed after hearing about the study. 6 out of the 11 participants had previously worked with one author as a peer (1), subordinates (3), or managers (2). Participants were uncompensated volunteers to avoid conflict of interest with their colleague (the interviewer).

We tried to go beyond our professional network, but recruitment was challenging. We contacted TH analysts from other agencies (DOE and DOD) and received one response. We ultimately excluded the other organizations from the study and focused on the one to which we had access.

Subject Demographics: The distribution of subjects by organization and by position is in Table 1.[1] Some subjects had operated on teams in multiple organizations within the last three years so the count in Table 1 exceeds 11. As precise titles could de-anonymize subjects, subjects are mapped to three general job roles: Leadership, Team Lead, and Analyst.

Table 1: Subject breakdown by organization and position.

Organizations	# Subjs.
DHS Organization # 1	10
DHS Organization # 2	3

Position	# Subjs.
Leadership	4
Team leads	4
Analysts	3

Table 2: Summary of interview protocol. The first column is the topic and the number of questions for that topic.

Topic (#)	Example questions
Demographics (2)	• How many missions have you been on? • How many found an adversary?
TH process (11)	• Draw your team's TH process. • What parts are problematic? • In what ways do you incorporate hypotheses into the process? • (Critical incidents) Tell us about 1-2 missions where the process failed?
New members (6)	• How long does it take a new team member to become productive? • What is a good measure of a member's expertise?

3.5 Interview Procedure

Interview protocol creation: The main body of our instrument focused on the knowledge gaps identified in §2.4: eliciting the TH process and understanding its challenges. Guided by previous TH literature, we asked about analysis frameworks and process automation [25, 37, 94]. Based on the frequency of personnel turnover, we also had a line of questions about incorporating new members.

Interview protocol refinement: Following best practices [38], we conducted two mock interviews. The primary researcher (who has TH experience) was interviewed by the other researchers. The protocol was tuned for clarity. Transcripts from these mock interviews were not used in analysis.

After the mock interviews, we divided the real interviews into two stages. In the first stage, we held two interviews to pilot our protocol, one with a CGCYBER team lead and one with a CISA team lead. After these interviews we reviewed the transcripts, assessed validity, and made changes as needed. Following this pilot, we added 2 questions and 3 follow-up questions, and re-worded one follow-up question. Over the course of the entire study, 92% of the main questions (22/24) were held constant. As there was little change in the instrument, our results include the 2 pilot interviews.

The final semi-structured interview protocol is summarized in Table 2 and detailed in the technical report (Appendix A). Interviews were conducted by one author, who had security clearance (for national security, §3.6) and prior TH experience (so that they could ask domain-appropriate follow-up questions). Interviews lasted 1 hour and used Microsoft Teams.

[1] Technical report has subjects' experience in years and # of missions.

3.6 Ethics and National Security

This study was approved by our institution's Internal Review Board (IRB). A signed consent form was collected from participants before their interviews.

The studied organizations have small TH communities so care was taken to anonymize subjects. Specifically, we removed identifying information from quotes, and we describe and quote subjects in terms of generic job roles (Table 1).

Due to the sensitive information being discussed, we took measures to ensure that no classified information nor data from customer sites was collected. The interviewer reminded each subject of the unclassified nature of the research. All audio was reviewed for sensitive information before being sent to the transcription service. The transcript was then reviewed again by the research team before being used for analysis.

3.7 Data Analysis

We used three types of analysis on the resulting transcripts. Process coding and inductive process discovery were used for RQ1. Thematic coding was used for RQ1 and RQ2. All codebooks are in the technical report (Appendix A).

3.7.1 RQ1: Process Identification

First, process coding was done as described in Saldaña [80]. In this method, a pre-existing process is represented in a codebook used to analyze (code) a transcript. We created a codebook from TaHiTI using their descriptions of hunt triggers. We likewise created a codebook for the Trent model. The subjects' descriptions of their TH process were then coded against these process codebooks to assess fit. [2]

In our results (§4.1), neither TaHiTI nor Trent was a good fit, so we induced a TH process model from the interview data. First, all nodes from all subject diagrams were placed into one interconnected diagram. Similar nodes were combined. When possible, more precise nodes took precedence over general nodes. If a node was only on one diagram and not mentioned in multiple subjects' interviews, that node was removed.

As this process was relatively objective, a single analyst combined nodes and made the diagram. Another analyst iteratively reviewed the resulting process model.

3.7.2 RQ2: Shortcoming Identification

Following process coding and diagram discovery, further analysis was conducted on the transcripts. Since the interview protocol contained primarily focused questions, the interviews proceeded in a readable linear fashion. This allowed one analyst to re-code the transcripts using thematic coding as described in Guest *et al.* [50]. Memos were written by the primary analyst. This researcher arranged the 636 memos into themes. Codes were iteratively refined, ultimately yielding 57 codes under 10 topics.A second analyst assessed reliability, using the codebook and excerpts from 6 of the 11 transcripts. For completeness, excerpts were selected from each set of Question-Answer in the interview protocol. The measured agreement was fairly high (Cohen's $\kappa = 0.82$).

Completeness of results: Saturation was measured after all 11 interviews were complete. We measured saturation following Guest *et al.* [49], by measuring the number of (cumulative) new codes appearing in each interview. We found saturation after seven subjects, with no new codes being observed in the last five interviews. All organizations and all position categories had been represented at that point, indicating substantial homogeneity. Codes on a per-interview basis were also charted. We found that each interview covered many topics ($\geq$30 codes per interview).

After analysis was complete we performed member checking [33] by circulating our results to two experienced subjects (one team lead and one member of leadership). They felt properly anonymized and that their TH data was represented.

3.8 Limitations and Threats to Validity

Like many qualitative studies, our primary limitations are the sample size (11 subjects) and the reliance on self-report.

- *N=11*: Guest *et al.* argue as few as six interviews can suffice if the population is homogeneous and the data collected is specific [49]. We believe this is the case here. Samples of ∼10 are common in interview studies [41, 53, 75, 79]. Also note our sample is relatively large compared to the population, which is ∼5,000 at the studied organizations.
- *Self-report:* For validity, we followed best practices in interview instrument creation [38] including conducting 2 internal mock interviews and slightly modifying the instrument after the first 2 subject interviews. We performed member checking [33] by circulating our results to two experienced subjects (one team lead and one member of leadership). They both felt properly represented. Unfortunately, triangulating against public documents was not possible, as those from TH teams lack process information [18] and internal documents are confidential.

We note two additional threats. First, the analysis was done primarily by one researcher. Bias is mitigated by measurements of inter-rater agreement on a subset of the data. Second, our study may not generalize, *e.g.,* to the rest of the government or to private sector teams. Government teams and agencies differ by mission, training, etc. The differences between the TaHiTI process and the Trent model suggest differences between public and private sector teams.

[2] The TH process description was the first part of the interview. The TaHiTI and Trent codebooks were only used on this part of the interview. It was rare for a subject to mention a new component of their process unprompted after their process description.

4 Results

4.1 RQ1: TH Processes of DHS Teams?

The process described by Government practitioners differed from the TaHiTI and Trent processes. We induce a unified process model capturing the participants' TH processes, comprising 7 stages and 25 distinct activities (Figure 2).

Each subject provided two types of data to indicate the process they used. First, at the beginning of their interview, 10 out of 11 subjects provided a process diagram. (The remaining subject did not feel sufficiently familiar with the process.) Second, in the remainder of the interview they described their team's process. We used process coding to compare our subjects' TH processes to the TaHiTI [94] and Trent [92] models. Our observed TH processes did not match (§4.1.1), and we describe the induced model in §4.1.2.

4.1.1 TH Process Comparison to Related Work

We coded processes against the TaHiTI and Trent models.

TaHiTI: We found TaHiTI a poor match to our subjects' processes. 88% of the TaHiTI codes matched to activities described by the subjects. However, every subject described at least one activity that occurred in addition to the TaHiTI process, including Baselining (8 subjects), Sensor Placement (8 subjects), Team Arrival (5 subjects), and Customer Meetings (5 subjects). We suggest two reasons for this poor fit:

(1) Internal Team Assumed: Some tasks that are only required for external organizations were omitted by TaHiTI. For example, when describing their process 8 subjects were mentioned placing sensors on site prior to the team's arrival. TaHiTI discusses data sources and ensuring data availability, but its concerns are different than our subjects' concerns about sensor placement. Sensor placement is not discussed by TaHiTI but was important to our subjects because their organizations operate on external third-party networks.

(2) Only Hypothesis Hunting Described: Some tasks described by subjects are more important for data-driven hunting than hypothesis-driven hunting. For example, baselining was described by subjects as a 1-3 day process in which the TH team, deployed on an unfamiliar network, takes time to document typical network behaviors. Subjects indicated that baselining was especially important for filtering out false positives and for behavior analysis. However, since TaHiTI describes hypothesis-based hunting, it does not cover data-driven hunting techniques like behavior analysis.

Trent: We found Trent a poor match to our subjects' processes because of the unit of analysis. The Trent model was designed to distinguish between cognitive tasks. Our subjects more often drew administrative or temporal distinctions between tasks. For example, the Trent model describes four types of analysis: Network, Host, Malware, and Forensic. Our subjects instead described their analysis work in terms of two processes, a manual behavior-driven loop and an automatic alert-driven loop. Both loops involved host and network analysis — although these are distinct cognitive tasks, most subjects did not perceive them as distinct process components. We observed the same dynamic with baselining — Trent groups baselining activities with similar activities in other phases of a hunt, while our subjects emphasized the importance of a separate task for baselining distinct from the analysis step of the process. Numerically, 21% of the process components mentioned by our subjects did not match any of Trent code.

Incorporation of Different Frameworks: Each subject was specifically asked about their use of the hypothesis method and three other models observed in TH descriptions across academia and the private sector (see §2.3). In both organizations, Mitre's ATT&CK framework was the most used, although its role differed across subjects.

4.1.2 Description of the Induced TH Process

We concluded that our subjects' TH processes differed from the available literature. We induced a unified process model following the method given in §3.7.1. Figure 2 shows the high-level result, with activities grouped into seven stages based on our judgment. A detailed version is in the technical report (Appendix A). This section walks through Figure 2.

(1) Begin Hunt: Subjects described two ways to begin a TH mission.[3] (1) In a *proactive* mission, the customer has no suspicion of adversarial activity but still wants a TH team to check. One subject said: "*Some of our [missions]...people just say: 'Well, we're interested in having you come see if anyone's here'*." (2) In a *triggered* mission, there is no specific indication of compromise but the customer believes an adversary may be undetected on their network and requests a TH team. In one subject's words: "*Maybe you get...intelligence that says, '...a system that belongs to you may be communicating to a malicious command and control infrastructure'*."

(2) Mission Planning: In the mission planning task, the TH team coordinates with the customer. A leader subject said: "*I think [planning is] the most critical piece...This is really what I think creates a useful engagement.*" This process typically starts with a customer's request, then a survey of their system and network admins, and then TH-customer meetings (*e.g.*, "*multiple technical phone calls...to discuss any details*"). Subjects described three components of mission planning: scoping, hypothesis creation, and mission plan creation.

Scoping is when the TH team decides on what parts of the network will be included in the hunt mission. Organizations may constrain the hunt in ways such as parts of their network,

[3] Subjects used different names for an individual hunting operation. We use the term *mission*. Third-party TH teams consider a mission as one engagement with a specific customer over a few weeks. In contrast, internal TH teams hunt continuously, with few-day missions by hypothesis.

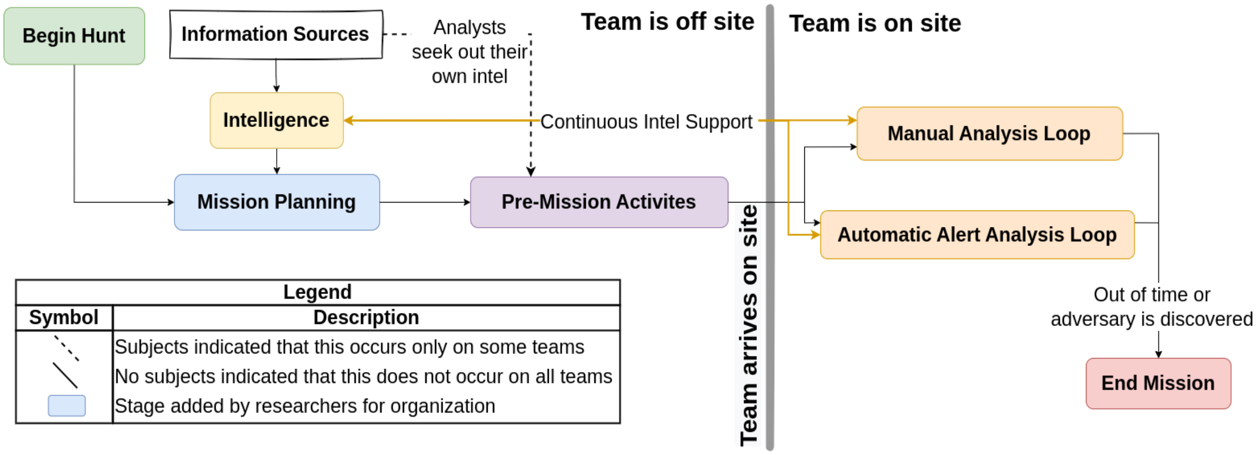
Begin Hunt
Information Sources
Analysts seek out their own intel
Team is off site
Team is on site
Intelligence
Continuous Intel Support
Manual Analysis Loop
Mission Planning
Pre-Mission Activites
Automatic Alert Analysis Loop
Team arrives on site
Out of time or adversary is discovered
End Mission
Legend
Symbol
Description
Subjects indicated that this occurs only on some teams
No subjects indicated that this does not occur on all teams
Stage added by researchers for organization

Figure 2: Unified TH process induced from interview data. A detailed version appears in the technical report.

types of threats considered, and time limit. In one subject's words: "*We got to figure out what kind of environment we're going to be working in, how many endpoints, how many users, inventory, ...how much data... stuff like that. We have a scoping questionnaire we send to the partner.*" Internal hunt teams often have a fixed scope of operation.

During *hypothesis creation*, a TH team generates a hypothesis to test in the hunt. Some teams do this during planning; others create hypotheses after deployment; others do not use hypotheses. Of the three studied organizations, two formally document hypotheses in the mission plan or elsewhere.

The *mission plan* documents the structure of the mission. It could include a timeline with associated deliverables, specific TH objectives, and 1-3 hypotheses if any. Not every team created a mission plan document. It may be unnecessary for internal teams as their hunt missions tend to be shorter.

Subjects in leadership positions often mentioned the importance of objectives in the mission plan, *e.g.*, "*The really important thing for TH is you go in with specific objectives. You're not just trying to find all bad activity.*"

Plans and objectives are not always communicated to the whole team. When asked if their team used a mission plan, one experienced analyst said: "*Yes, but I didn't have access — I was just a lowly [low rank] at the time. It was the team lead who ... would come up with it.*"

(3) Collect Intelligence: Subjects described 3 intelligence collection tasks. (1) When a specific trigger exists, intelligence is tailored around that trigger. (2) On proactive missions, current attacker trends are used: "*If there's a current prevalent exploit that's being used, like that you're seeing in the news... we're gonna go ahead and look for those.*" (3) Regardless if the hunt is proactive or triggered, teams will often collect and upload additional Indicators of Compromise (IOCs) to achieve increased coverage: "*we'll...load IOCs from all the threat intelligence providers.*"

Subjects provided many example intelligence sources but 4 sources were recurring: (1) A distinct intelligence team supporting the TH team; (2) Publicly Available Information (PAI) like social media or news feeds; (3) Open-Source Repositories like InQuest's Indicators of Compromise (IOC) database [23]; and (4) Classified or Subscription Feeds.

(4) Pre-Mission Activities: Before a TH team begins a mission, they fine-tune their tools using the Intelligence and the mission scope, hypothesis, and plan. One subject described typical pre-mission activities and how these can vary between internal and external hunts:

> "*So you plan the mission...dates [and] goals... And then you'll push that into developing your specific tools and analytics. If you're hunting yourself [i.e., internally], you probably already have most of your sensors and tools in place ... If you don't have tools or specific analytics to cover what you're going after, then you need to develop or buy or get those plugged in.*"

(5/6) Manual and Automatic Analysis Loops: At this point, the TH team deploys to the customer's site and spends 1–4 weeks hunting. Most of the TH time is spent here.

Subjects described two different modes of analysis. Both modes included cyclical tasks so we term them "loops". In the Manual Analysis Loop, TH team members enrich their baseline and perform manual analysis of the collected data, looking for potentially malicious behaviors. In the Automatic Alert Analysis Loop, members triage sensor notifications of possible IOCs. These loops continue concurrently and members are assigned to them at the team lead's discretion. A team lead summarized: "*Those two processes...are our constant back and forth over a couple weeks that we're doing on-site until we can either find something or not.*"

One subject described these loops ("paths"): "*So the first path is your alerting path [Note: the automated loop]. Those are your indicators and signature based detections ... And then the second path [Note: the manual loop] I see kind of starts with understanding the environment which is some baseline analysis...that path then feeds into your two like core detections, which are ... your behavioral analysis ... and then you* *also have anomaly based analysis.*”

(7) End Mission: The goal of a hunt mission is to detect adversaries, if they exist on the network. If adversary activity is observed, the hunt is over and a response is necessary. One subject said: “*If we find commodity malware on the computer, it may not require a lot of resources...If it’s something bigger...we’re probably going to shift into IR mode.*”

At the end of the mission, TH teams report their findings and provide recommendations: “*We come back to those hypothesis in our final report, too....[the report] goes to the entity and back to our leadership saying, ‘We tried to see if your exchange server was compromised by doing X, Y, and Z. Here’s some things we investigated as a part of that. Here’s our conclusion or what that led us to believe.*’”

4.1.3 Process management

Process Creation: Process creation differed across organizations. One organization’s process was devised by a group of team leads. At another organization, the TH process was created by 1-2 people in leadership positions and was deliberately imitative of other government TH teams. This imitation lets their TH teams interoperate with other organizations, enabling personnel sharing in the event of a national cyberattack.

Process Changes: At both organizations, the TH process changes frequently, and changes can be proposed by any experienced personnel: “*everyone who had prior experiences or is now part of the [team] and been on a mission...can submit edits.*” Subjects describe the process changes positively, in terms of continuous improvement — “*We’re constantly looking for how we can improve... So basically after every mission... we do a hot wash and we say: ‘Okay, what could have gone better?’*”. One team lead did warn that process change in the middle of a mission can be problematic: “*If you change the tactics too often, you are going to tire out your analysts ... You’re gonna tire out yourself and you’re gonna get confused when you go to write the final report.*” They implement changes between missions instead.

4.2 RQ2: Process shortcomings and solutions?

Table 3 shows process issues, proposed solutions, and open questions. Common process challenges are (1) determining expertise of team members (to assign appropriate tasks); and (2) Improving automation.

This section describes TH process challenges noted by ≥ 3 subjects across both organizations, and possible solutions where available. Shortcomings are ordered by frequency. To avoid biasing to a single subject’s “pet peeves”, in this section all subjects are quoted at least twice and no more than 5 times.

4.2.1 Identifying expertise

Problems: TH team leaders want to assign members different tasks depending on the members’ expertise. Many indicators of expertise exist — three common indicators are external certifications, internal certifications, and training. Subjects seemed to dislike these. Instead, they used *ad hoc* indicators such as experience and time spent working off-hours.

(1) Certifications: Six subjects did not believe certifications were a good measure of expertise. One subject was skeptical that certifications measured even baseline knowledge: “*It’s really hard to define what a baseline of cybersecurity understanding is ... You can’t really say it’s having a certain certification because there’s plenty of people that have certifications that don’t know what they’re talking about.*”

(2) Internal Certification: One organization maintains an internal certification as mentioned by five subjects. One subject stated that it helped the team be interoperable with other government TH teams. Three subjects voiced dissatisfaction with the certification stating it was irrelevant or poorly implemented: “ *I have some mixed feelings on [the certification]...It doesn’t ...apply to us all that well and the proficiency needed to complete it isn’t ...much either.*”

(3) Training: Two subjects, both inexperienced, spoke favorably of training as an indicator of expertise: “*I think it’s just training experience, time with the tools, time with the knowledge base — I think training has a big part of it though. And I think just hands-on is pretty key as well, which the training could help with.* ” Two experienced subjects were more cautious. For example, one said: “*There is a mild or weak correlation between number of training courses and analyst success*”. Four subjects spoke unfavorably of training as an indicator of expertise. A team lead said “*having a [certain] course under your belt or any other course, usually doesn’t give me immediate confidence.*”. One analyst strongly critiqued their required training courses: “*We have [an agency] requirement... just a massive waste of time. <laugh>*” However, this analyst recognized that some trainings can help: “*I’ve given the team several trainings [about tools]...little exercises to help people get up to speed.*”

Solutions: Developing indicators of expertise remains an open question. Subjects had some suggestions, which we list next, though each idea was opposed by ≥1 subject.

(1) Time with tools: The importance of time with the tools was a theme that often accompanied discussions about training, as an indicator of expertise or a tactic to integrate newer members. When subjects discussed the value of training, they often spoke in terms of whether it let them improve their skills with their tools. For example, one subject said: “*looking at the [training]... it’s a good experience but... they don’t use the tools we currently use...It doesn’t provide the training that we need per se, for the tools that we use.*”

(2) Working on personal time: Working off-hours was viewed more positively than training or certifications. Five

Table 3: Process challenges and proposed solutions noted by subjects. Open problems have no consensus solution or occur when subjects verbalized difficulty or uncertainty. Only challenges noted by ≥ 3 subjects and observed in both orgs. are included.

Observed process challenge	# subjects (# orgs)	Proposed Solutions	Open Problems
Poor measures of expertise (§4.2.1)	6 (2)	Make training more tool oriented	Determining TH expertise?
Insufficient automation (§4.2.2)	6 (2)	Automate baselining & set-up tasks	Will automation hinder the analysts?
Teams lack needed data (§4.2.3)	5 (2)	Deploy sensors ahead of time	What data to collect?
Inappropriate process detail (§4.2.4)	4 (2)	Give team leads flexibility	*None mentioned by subjects*
High turnover (§4.2.5)	3 (2)	Pair new and experienced members; Enhance process documentation	*None mentioned by subjects*

subjects said doing cybersecurity activities on personal time was a good metric for potential or expertise. Subjects made comments like: "*I can't really put my finger on a single thing that...explains our expertise. I would say, a big part of it is just being willing to kind of play around with the tools and even in their own time. I mean, we have a lot of downtime between missions where we do trainings, but I'm thinking more, even within that time-frame, they have a lot of free time where they can kind of play around with the tools or do you know, capture the flags ...that almost has been a bigger indicator.*" We note that not all analysts will be able to devote free time and that making this an indicator of expertise could have undesired effects such as marginalizing certain otherwise high performing analysts as has happened in similar cyber security fields [46].

(3) Experience: Ten subjects mentioned experience (*e.g.*, the number of missions) in the context of expertise. Five subjects were generally positive about experience as an indicator of expertise, while four demurred. One subject said: "*I think experience is a key metric. ...most of the guys that we got that had already done threat hunting [before joining our organization] are our best analysts*". A common counter-argument was: "*Experience does not equal quality...just because someone has experience doesn't mean they're good at their job.*"

Subjects who affirmed experience as an indicator of expertise often did so hesitantly, recognizing it as a heuristic. For example, the positive subject quoted earlier also said "*You...have people who've been on a hundred missions and they just don't have the aptitude or the thought ability. You...have people who have never been on a mission who have the aptitude and who are gonna outperform people who have been on 20 missions. ... I think it's a little hard to answer that question directly, but I would say the biggest <hesitates> there is a correlation between number of missions and analyst success.*"

4.2.2 Improving Automation

Problems: Automation is often emphasized as a way to reduce costs or to discover adversaries more quickly. The subjects indicated their automation was insufficient or ineffective.

(1) Insufficient Automation: One issue was that both organizations had little automation. When subjects were asked how much of their process was automated, answers ranged from "*None of it's automated*" to "*25 – 35%*".

Six subjects indicated dis-satisfaction with the current level of automation. They made statements like "*that's a big point that could be improved*". Two additional subjects indicated that the process was all or mostly manual, making statements like: "*It's still pretty manual.*"

Seven subjects described hindrances to more automation. Two subjects mentioned that analysts had insufficient time to generate automation. Subjects made comments like "*Mainly just lack of time development time. We've just been very busy.*" Two subjects described a lack of personnel. For example one subject responded: "*Probably just personnel gaps.*" Two subjects mentioned that the team did not have the necessary knowledge to automate tasks that should be automated. When asked why more had not been automated one of these two subjects said: "*because nobody understands [tool]*".

Two subjects hesitated about further automation. One team lead did not want the automation to become a crutch. Another subject (in a leadership role) had a philosophical concern:

> "*Hunting...should start where automation stops. That's...the whole premise of hunting...the sophisticated actor already bypassed all that [automation] and now you have to apply manual techniques to...find them... It's good to automate [repetitive] things, maybe like deploying kit. But I think most of the analytical work should...rely on human factor.*"

(2) Ineffective Automation: A second issue with automation is its effectiveness. Figure 2 includes two analysis loops: manual and automated. Out of the subjects that had detected adversaries on hunt missions (7/11), none of their examples described activity found by the automated alert loop. One team lead said: " *[The adversary] was almost exclusively found by analysts observing the data for living off the land techniques or new zero days.*[4] *I would say that we found at least three or four zero days, which, couldn't have been detected by [automation or intelligence reports].*" A member of leadership said: " *[What] we find are mostly analytical kind of behavioral activity. Like an analyst spots.*"

Solutions: Five subjects suggested opportunities for greater automation. Two subjects said baselining took too long and should be automated. One said: "*Have the baseline step be an*

[4] "Living off the land": using the victim's own tools instead of downloading external tools. "Zero day" vulnerabilities: unknown to network defenders.

automatic step. ... where baseline is also is almost automatic. ... then you can kind of shift the window baseline out from two to three days...to a five minute thing." Two other subjects discussed equipment preparation tasks. One suggested: "*It would be cool to automate the indicator portion, other than us having to manually feed in that stuff.*" Subjects also suggested automating endpoint log collection and automating the reporting of important artifacts and their contexts to the team.

4.2.3 Improving Data Collection

Problems: A successful threat hunt needs timely and sufficient data. Subjects noted challenges in both dimensions.

(1) Delayed data collection: Many teams would only begin collecting data when the team arrived on site. This created an issue because the team was forced to start hunting before much baseline data had been collected. Without data, threat hunting is virtually impossible or as one experienced team lead said: "*you can't find anything without data*". Seven subjects either mention this issue or said that they have learned to deploy sensors ahead of time to combat it. One team lead said: "*We did a threat hunting engagement...for about two weeks, but we plugged sensors in on Day One of those two weeks. When you do that, you don't have a baseline of what even one work week looks like much less a couple...we ... didn't have the baseline to actually do real analysis on the network.*"

(2) Un-prioritized Data Collection: Some TH teams collected too much data or the wrong type of data. This extraneous data obscures potentially relevant data. One experienced member of leadership said: "*I know with [former TH organization]...we would ask for a ton of stuff. Give us this, and that...and for no reason, right? ... Just because you have all the data does not necessarily mean that you're going to be more effective doing your hunt. I think it's the opposite and that's kind of the mindset that needs to change.*"

Another subject, a team lead, gave an example of a mission where so much data was collected that the servers containing the data stopped behaving properly. The subject reported that in this environment: "*if you tried to follow the checklist you ended up getting really confusing results because the logs that were coming in were non-deterministic ... trying to stick to [the checklist] was counterproductive because if you tried to stick to it, the data made less sense.*"

Solutions: Five subjects had been on teams that had experienced the issue of lacking data when the team began analysis. These subjects described a solution that worked for their teams: Weeks before the TH teams' arrival, a predeployment team would install sensors on the network. One subject said: "*I'll bring up a problem that we had. <laugh>our first one that we've worked through, which was a huge improvement, was the prep work in advance, right? [Before] we just set sensors the day that we arrived and that was terrible.*"

The issue of unhelpful data being collected seemed to be difficult. No subjects mentioned having solved this issue. One experienced member of leadership suggested more precise scoping as a possible solution. They said: "*We go into these environments, you see a lot of guest networks, or IOT devices, like cameras and security badges and stuff. It's stuff that we don't necessarily need to monitor ...and then it just fills up our sensors ...[queries] take forever to get results. Let's concentrate on one thing, but do more in depth work. Concentrate on quality, I guess, versus quantity.*"

4.2.4 Improving Process Documentation

Subjects expressed opposing views on their team's process documentation, as well as what would constitute good process documentation. We therefore do not present this theme in terms of problem-solution, but rather in terms of these opposing views. Conflicts in these views are reminiscent of the "contradictions" observed in SOCs by Sundaramurthy *et al.* [89]. These views may be influenced by the level of detail in process documentation on each team. One organization had a detailed checklist. Another had a similarly detailed process but allowed team leads to adapt it per deployment.

View–More detailed process documentation: Four subjects requested or supported more detailed process documentation. "*I think having the [detailed process documentation] will help [new members] be more effective, faster, because it will give them a guideline of what to do. So instead of sitting there not knowing what the first step to take, at least they know a general network to look in. So instead of sitting there: 'I don't even know where to start.' It's: 'I'm looking for connections on odd ports. That's my first thing. And that's how I'm gonna learn to build my first queries.'* " More documentation and process rigidity was considered helpful for newer members. One analyst described the creation of a spreadsheet documenting the task that had to be completed on a TH mission: "*We didn't really know what we were looking for before spreadsheet. ... I took the idea of the spreadsheet by looking at Mitre ATT&CK and seeing what we could look for. ... so it helped [new members] focus on each task at a time.*"

View–Less detailed process documentation: In contrast, four subjects spoke in favor of less detailed process documentation. One subject said their team's process was too detailed: "*For the current experience level of [my team], I would like it to see it a bit more vague... [subject gave an example of a VPN checklist item for a network without VPN] ...It's too specific, you know? ...[We should] make that more abstract and say like, 'Hey, why don't [you] just look for any remote access software?'*". One organization addresses this conflict by giving team leads autonomy: "*[Team leads] are responsible for their mission...They are given leeway to adjust as things happen.*" When asked how often team leads adjust the process, this subject said: "*I'd say 25 to 50% of the time*".

In support of this view, one team lead gave an example of a mission where the process felt over-specified and switching to a more abstract process improved their performance. Another subject in leadership spoke to the importance of flexibility: "*I dislike making rigid guidelines ...I think the goals of each engagement may be different based on the partner, the threat actor, the geopolitics ...I think making it too strict makes it so that we lose some of the flexibility to do the missions that have the most impact or that we can't meet the goals we want based on a checklist that wasn't developed for them.*"

4.2.5 Turnover

Problem: Three subjects mentioned the negative effects of turnover. One subject said: "*our biggest problem right now is turnover ... [at] any given time, you maybe have a quarter to a third of your team, [that] has been on more than one mission*". They mentioned a case where their entire team was replaced with new members.

Solutions: Two possible solutions were offered by subjects: pairing members and improving process documentation.

(1) Pairing Members: Six subjects recommended pairing new members with expert members to improve the integration of newer members. One less experienced subject found it quite helpful, saying "*[Observing more experienced members] was nice because I was basically able to look over the shoulder over all the missions that have happened so far and see where I fit in the puzzle piece of Threat Hunting.*".

Two more experienced subjects agreed that pairing members can help, but voiced warnings. One spoke about short- vs. long-term payoffs: "*It's a balance between, do we need to succeed in this mission or do we need train our junior analysts to succeed in the next mission?*" The other emphasized that pairing requires learner engagement: "*If the new member isn't particularly motivated then they're just gonna be staring at a screen and not learning anything ...it would be nice to ...have ...a side saddle process...[where] let's both do it at the same time separately [and] you show me how to do it.*"

(2) Process Documentation: Five subjects described good process documentation as important for assisting new members. Some subjects felt that process documentation helps a team meet a minimum standard, *e.g.,* "*I think defining...baselines...really well so you have the minimum standards and a clear task list... What [are] the 10 steps that you absolutely will do before going deeper? And if you have that well defined and built into your tools with automations when possible, that makes it...easier for people to come on board and get up to speed.*" Another subject said: "*Usually the less experienced ones are better at following the process, mostly because they don't know any better to do anything else.*"

5 Discussion & Future Work

Our work has two audiences: operations teams and researchers. We discuss TH among other cybersecurity operations (§5.1), then share implications of our study for TH teams (§5.2) and opportunities for future research (§5.3).

5.1 Threat Hunt and Other CyberOps

As discussed in §2, TH work is adjacent to Incident Response (IR) and Security Operation Centers (SOCs). The experiences of our DHS TH team member subjects thus echo, yet are distinct, from the experiences of IR and SOC teams.

The DHS TH teams described many challenges shared by IR teams. Our subjects said that measuring TH team effectiveness is difficult due to lack of feedback; IR teams have a similar issue [93]. IR analysts express similar concerns about automation hindering analysis capability [73]. Data access is an issue for both TH and IR teams [48]. However, IR teams differ from our TH subjects in the data they have about the adversary. One reason why models such as ATT&CK and Kill Chain are popular is that they show IR teams the "next step" when investigating an incident. For TH teams (and SOCs), there is no confirmed incident, so a TH team must investigate everything, hoping a clue was dropped along the kill chain.

With respect to SOCs, the studied TH teams operate as a third-party "check" on the SOCs' networks. Our subjects thus undertake similar activities, but with different information. The DHS TH teams get intelligence from other government entities, much of which cannot be shared with private sector SOCs. The DHS TH teams also set up their own sensors to check for blind spots in the SOCs' surveillance. Beyond these informational strengths lie weaknesses: DHS TH teams struggle to baseline because they are not on the network for long. This may be why some subjects and other TH researchers that believe that TH can never be fully automated [31, 66] — automation requires a baseline that third-party hunters lack.

5.2 Recommendations for Threat Hunt Teams

In §4.2 we presented challenges and potential solutions. Now we synthesize our observations into three recommendations.

(1) Improve Planning. Although many subjects focused on the TH activities *after* deploying to a customer site, about half of our unified TH process (Figure 2) occurs before deployment. We were surprised by the range of planning activities and formality reported by subjects. We recommend that all TH teams create mission planning documents **and** share the objectives with everyone on the team. Following the advice of one subject: "*plan the mission...your high-level goals...and push that into...specific tools and analytics.*"

(2) Revisit the Automated Alert Loop. In §4.1.2 we described the two analysis loops applied by the TH teams we studied (see Figure 2). However, in §4.2.2 our subjects reported that their automated alert loops rarely find adversary activity. There are many possible causes, including redundant automation (*e.g.,* already applied by SOC teams), human error (*e.g.,* important alert is forgotten and not uploaded to tool) and absent automation (*e.g.,* zero-day exploits). We recommend either making the automated alert loop more effective or reducing the resources devoted to it.

The ineffectiveness of automated alerts might be a flaw in threat intelligence. One analyst, describing adversary detection on a previous mission, was surprised that there was no alert for the activity. They said: "*I don't think there was an alert on it. Which is odd thinking about it now...you definitely would think that [tool] would contain some...but I don't think there was.*" Eight subjects claimed that their intelligence component supplies the team with at least some of the alerting rules used during a hunt. Perhaps TH teams should revisit the trust they place in these turnkey automated rules.

If TH teams do not revisit their automated alert strategy, then they might re-evaluate how many resources they invest in it. For example, one subject estimated that intelligence/alerting only accounts for "*maybe a tenth*" of adversary detections. If the amount of resources put into this analytic loop exceeds 10%, this could indicate an inefficiency.

(3) Formalize apprenticeship. Many subjects voiced concern about the quality of existing TH training and certifications §4.2.1. Some teams addressed this by pairing new and experienced members (§4.2.5). These teams appear to be reinventing the concept of *apprenticeship*, which is an educational strategy for disciplines that are more art than science [1]. There is much literature on the virtues and shortcomings of apprenticeship [35, 77] and apprenticeship is currently being used with success in cybersecurity [85, 87]. Since subjects seem to agree that pairing is a good way to integrate new members, teams may benefit from a more robust apprenticeship program. According to subjects, this should include providing time while on mission, since hunting with trainees takes longer. We recommend ensuring trainees not just watch others hunt, but rather interact with data and tools under supervision (the "side saddle" process suggested by one subject).

5.3 Future Work

Threat Hunt is a young cybersecurity discipline. Our exploratory study suggests many beneficial directions of study.

Tailor automation to needs: We reported subjects' suggestions for automation in §4.2.2. The most popular suggestions were automating baselining and recurring equipment set-up tasks. In contrast, most of the automation in the academic literature focuses on automating the entire TH process [52, 56, 67, 96]. This class of automation may address the current shortcomings of the automated analysis loop. However, it does not address the needs verbalized by our subjects.

Does automation hinder analysts? In §4.2.2, one team lead and one subject in a leadership role expressed concerns that automation would lead analysts to not think for themselves. Measuring whether, and under what conditions, automation reduces team effectiveness would provide useful data for TH decision-makers.

Process formalism or flexibility? In §4.2.4, subjects request more detailed TH process documentation. However, in that section, an example is given of a process that was too precise, becoming a hindrance to the team. A second subject indicates that they wish the process were more vague. Some teams deal with this tension by allowing team leads to make process changes. Not all organizations require their team leads to have TH experience so not all team leads will be capable of making these decisions. It is possible that this tension is inherent to a creative and open-ended activity, but tracking and measuring it may be helpful.

Process evaluation: Subjects did not agree on metrics for measuring TH team effectiveness. The goal of TH is to reduce adversary dwell time [94]. However, this metric may not be applicable to the TH teams we studied because they act in a third-party capacity and do not necessarily revisit organizations. Additionally, organizations may be compromised so infrequently that measuring dwell times does not provide a TH team with actionable feedback.

Other metric recommendations include security improvement [45], and risk reduction [51]. These metrics measure whether a TH team added value by identifying vulnerabilities or security blind spots. An objective measure of this kind of effectiveness may help evaluate the quality of a TH process. A final proposal is Bianco's Hunt Maturity Model [31], which measures a team's capabilities rather than its products. None of these metrics has been systematically evaluated in practice.

6 Conclusion

Many organizations have recently adopted Threat Hunting as a way to detect adversaries who have infiltrated their networks undetected. There is little academic literature on Threat Hunting processes, and no description of the Threat Hunting process as performed by US government hunt teams working on third-party networks. In this work, we provide the first description of the US government Threat Hunting process model as practiced by teams in the US Department of Homeland Security. We found that these teams have different processes than those reported in prior literature, differing from both private-sector internal hunt teams and other government teams. We provide a novel model of the Threat Hunt process, complemented with a set of open problems and possible solutions suggested by Threat Hunt practitioners. Much work remains: in the short term, these process recommendations can be implemented and tested; in the long term, experiments in this sensitive context remain an open challenge.

7 Acknowledgments

We acknowledge support from the US Coast Guard. We thank our research subjects for their participation. We thank G. Cramer for assistance in coding, and A. Kazerouni for feedback on the manuscript. We thank the CCS and USENIX reviewers for constructive criticism.

References

[1] The Cambridge Handbook of the Learning Sciences. Cambridge Handbooks in Psychology. Cambridge University Press, 2 edition, 2014. doi:10.1017/CBO9781139519526.

[2] Combat Mission Teams and Cyber Protection Teams Lacked Adequate Capabilities and Facilities to Perform Missions. Technical Report DODIG-2016-026, Inspector general of the US Department of Defense, November 2015.

[3] Gaining The Advantage, 2015. URL: https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/Gaining_the_Advantage_Cyber_Kill_Chain.pdf.

[4] Lessons to Learn from the OPM Breach, June 2015. URL: https://www.tenable.com/blog/lessons-to-learn-from-the-opm-breach.

[5] Lessons from the OPM breach, September 2016. URL: https://gcn.com/cybersecurity/2016/09/lessons-from-the-opm-breach/316728/.

[6] Cybersecurity Experts Hunting for Hackers, April 2017. URL: https://www.nationaldefensemagazine.org/articles/2017/4/3/cybersecurity-experts-hunting-for-hackers.

[7] National Cyber Strategy, September 2018. URL: https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf.

[8] The Cost of Cybercrime. Technical report, Ponemon Institute LLC and Accenture, 2019.

[9] Cost of a Data Breach Report 2020. Technical report, IBM Security, July 2020. URL: https://www.ibm.com/security/digital-assets/cost-data-breach-report/1Cost%20of%20a%20Data%20Breach%20Report%202020.pdf.

[10] Guide To Cyber Threat Hunting, 2020. URL: https://www.tylertech.com/services/ndiscovery/nDiscovery-Threat-Hunting.pdf.

[11] How High Employee Turnover Poses Increased Cyber Security Risk, November 2020. URL: https://copperbandtech.com/high-employee-turnover/.

[12] 6 U.S.C. §659, December 2021.

[13] Attack vs. Data: What You Need to Know About Threat Hunting | Rapid7 Blog, March 2021. URL: https://www.rapid7.com/blog/post/2021/03/25/attack-vs-data-what-you-need-to-know-about-threat-hunting/.

[14] Cyber Security Analyst Demographics And Statistics In The US, December 2021. URL: https://www.zippia.com/cyber-security-analyst-jobs/demographics/.

[15] Executive Order 14028: Improving the Nation's Cybersecurity, May 2021. URL: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/.

[16] CYBER 101: Hunt Forward Operations, November 2022. URL: https://www.cybercom.mil/Media/News/Article/3218642/cyber-101-hunt-forward-operations/https%3A%2F%2Fwww.cybercom.mil%2FMedia%2FNews%2FArticle%2F3218642%2Fcyber-101-hunt-forward-operations%2F.

[17] Falcon Overwatch: Managed & Proactive Threat Hunting | CrowdStrike, 2022. URL: https://www.crowdstrike.com/endpoint-security-products/falcon-overwatch-threat-hunting/.

[18] Malicious Cyber Actors Continue to Exploit Log4Shell in VMware Horizon Systems. Cybersecurity Advisory AA22-174A, Cybersecurity and Infastructure Security Agency, June 2022.

[19] Talos Incident Response, 2022. URL: https://talosintelligence.com/incident_response/hunting.

[20] Threat Hunting, 2022. URL: https://www.boozallen.com/expertise/cybersecurity/threat-hunting.html.

[21] The United States Government Manual, 2022. URL: https://www.govinfo.gov/content/pkg/GOVMAN-2022-12-31/xml/GOVMAN-2022-12-31.xml.

[22] X-Force Threat Intelligence Index 2022. Technical report, IBM Security, February 2022. URL: https://www.ibm.com/downloads/cas/ADLMYLAZ.

[23] InQuest Labs - IOCDB - InQuest.net, 2023. URL: https://labs.inquest.net.

[24] National Cyber Strategy, March 2023. URL: https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf.

[25] Anchit Agarwal, Himdweep Walia, and Himanshu Gupta. Cyber Security Model for Threat Hunting. In 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pages 1–8, September 2021. doi:10.1109/ICRITO51393.2021.9596199.

[26] Norah Alharbi. A Security Operation Center Maturity Model (SOC-MM) in the Context of Newly Emerging Cyber Threats. Ph.D., The Claremont Graduate University, United States – California. ISBN: 9798672151229.

[27] Rahaf Alkhadra, Joud Abuzaid, Mariam AlShammari, and Nazeeruddin Mohammad. Solar winds hack: In-depth analysis and countermeasures. In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pages 1–7. IEEE, 2021.

[28] Applebaum, Andy, Nickels, Katie, Schulz, Tim, Strom, Blake, and Wunder, John. Getting Started with ATT&CK, October 2019. URL: https://www.mitre.org/sites/default/files/2021-11/getting-started-with-attack-october-2019.pdf.

[29] Frederico Araujo, Dhilung Kirat, Xiaokui Shu, Teryl Taylor, and Jiyong Jang. Evidential cyber threat hunting, 2021. arXiv:2104.10319.

[30] James E Bartlett, Joe W Kotrlik, and Chadwick C Higgins. Organizational Research: Determining Appropriate Sample Size in Survey Research.

[31] David J Bianco. A Simple Hunting Maturity Model. URL: http://detect-respond.blogspot.com/2015/10/a-simple-hunting-maturity-model.html.

[32] David J Bianco. A Framework for Cyber Threat Hunting Part 1: The Pyramid of Pain, 2015. URL: https://www.threathunting.net/files/A%20Framework%20for%20Cyber%20Threat%20Hunting%20Part%201_%20The%20Pyramid%20of%20Pain%20_%20Sqrrl.pdf.

[33] Linda Birt, Suzanne Scott, Debbie Cavers, Christine Campbell, and Fiona Walter. Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? Qualitative Health Research, 26(13):1802–1811, November 2016. doi:10.1177/1049732316654870.

[34] Everett Bledsoe. How Often Do Military Families Move? Why They Move So Much?, January 2023. URL: https://www.thesoldiersproject.org/how-often-do-military-families-move/.

[35] Michaela Borg. The apprenticeship of observation. Elt Journal, 58:274–276, July 2004. doi:10.1093/elt/58.3.274.

[36] Derek E Brink. Quantifying the Value of Time in Cyber-Threat Detection and Response. Technical Report 15218, Aberdeen Group, January 2017.

[37] Jon R Bynum. Cyber Threat Hunting. Master's thesis, Utica College, 2019.

[38] Ronald J Chenail. Interviewing the Investigator: Strategies for Addressing Instrumentation and Researcher Bias Concerns in Qualitative Research. The Qualitative Report, 16(1):8, January 2011.

[39] Selina Y. Cho, Jassim Happa, and Sadie Creese. Capturing Tacit Knowledge in Security Operation Centers. IEEE Access, 8:42021–42041, 2020. doi:10.1109/ACCESS.2020.2976076.

[40] Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone. Computer Security Incident Handling Guide. Technical Report NIST SP 800-61r2, National Institute of Standards and Technology, August 2012. doi:10.6028/NIST.SP.800-61r2.

[41] Martina De Gramatica, Fabio Massacci, Woohyun Shim, Alessandra Tedeschi, and Julian Williams. IT Interdependence and the Economic Fairness of Cybersecurity Regulations for Civil Aviation. IEEE Security & Privacy, 13(5):52–61, September 2015. doi:10.1109/MSP.2015.98.

[42] Joan E Dodgson. Reflexivity in qualitative research. Journal of Human Lactation, 35(2):220–222, 2019.

[43] Charlette Donalds and Kweku-Muata Osei-Bryson. Cybersecurity compliance behavior: Exploring the influences of individual decision style and other antecedents. International Journal of Information Management, 51:102056, April 2020. doi:10.1016/j.ijinfomgt.2019.102056.

[44] Douglas Everson and Long Cheng. Network Attack Surface Simplification for Red and Blue Teams. In 2020 IEEE Secure Development (SecDev), pages 74–80, September 2020. doi:10.1109/SecDev45635.2020.00027.

[45] Paul Ewing and Devon Kerr. The Endgame Guide to Threat Hunting. URL: https://cyber-edge.com/resources/the-endgame-guide-to-threat-hunting/.

[46] Kelsey R Fulton, Samantha Katcher, Kevin Song, Marshini Chetty, Michelle L Mazurek, Chloé Messdaghi, and Daniel Votipka. Vulnerability discovery

for all: Experiences of marginalization in vulnerability discovery. In 32nd USENIX Security Symposium (USENIX Security). USENIX Association, 2023.

[47] Peng Gao, Fei Shao, Xiaoyuan Liu, Xusheng Xiao, Zheng Qin, Fengyuan Xu, Prateek Mittal, Sanjeev R. Kulkarni, and Dawn Song. Enabling Efficient Cyber Threat Hunting With Cyber Threat Intelligence. In 2021 IEEE 37th International Conference on Data Engineering (ICDE), pages 193–204, April 2021. ISSN: 2375-026X. doi:10.1109/ICDE51399.2021.00024.

[48] George Grispos, William Bradley Glisson, and Tim Storer. Security incident response criteria: A practitioner's perspective. CoRR, abs/1508.02526, 2015. arXiv:1508.02526.

[49] Greg Guest, Arwen Bunce, and Laura Johnson. How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. Field Methods, 18(1):59–82, February 2006. doi:10.1177/1525822X05279903.

[50] Greg Guest, Kathleen MacQueen, and Emily Namey. Applied Thematic Analysis. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States, 2012. doi:10.4135/9781483384436.

[51] Robert Gutzwiller, Josiah Dykstra, and Bryan Payne. Gaps and Opportunities in Situational Awareness for Cybersecurity. Digital Threats: Research and Practice, 1(3):1–6, September 2020. doi:10.1145/3384471.

[52] Antonio José Horta Neto and Anderson Fernandes Pereira dos Santos. Cyber Threat Hunting Through Automated Hypothesis and Multi-Criteria Decision Making. In 2020 IEEE International Conference on Big Data (Big Data), pages 1823–1830, December 2020. doi:10.1109/BigData50022.2020.9378213.

[53] Jan Huck and Frank Breitinger. Wake Up Digital Forensics' Community and Help Combat Ransomware. IEEE Security & Privacy, 20(4):61–70, July 2022. doi:10.1109/MSEC.2021.3137018.

[54] Pierre Jacobs, Alapan Arnab, and Barry Irwin. Classification of Security Operation Centers. In 2013 Information Security for South Africa, pages 1–7, Johannesburg, South Africa, August 2013. IEEE. doi:10.1109/ISSA.2013.6641054.

[55] Jason Chaffetz, Mark Meadows, and Will Hurd. The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation. Majority Staff Report, U.S. House of Representatives Committee on Oversight and Government Reform.

[56] Prakruthi Karuna, Erik Hemberg, Una-May O'Reilly, and Nick Rutar. Automating Cyber Threat Hunting Using NLP, Automated Query Generation, and Genetic Perturbation. arXiv:2104.11576 [cs], April 2021. arXiv: 2104.11576.

[57] Sean Michael Kerner. Lessons Learned from the OPM Breach eSecurity Planet, December 2017. URL: https://www.esecurityplanet.com/threats/lessons-learned-from-the-opm-breach/.

[58] Klaas-Jan Stol, Paul Ralph, and Brian Fitzgerald. Grounded Theory in Software Engineering Research: A Critical Review and Guidelines. In 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering, Florence, Italy, 2015. IEEE.

[59] Written Rob Lee and Robert M Lee. The Hunter Strikes Back: The SANS 2017 Threat Hunting Survey, 2017. URL: https://www.sans.org/webcasts/threat-hunting-modernizing-detection-operations-2017-threat-hunting-survey-results-1-103767/.

[60] Written Robert M Lee and Rob Lee. The Who, What, Where, When, Why and How of Effective Threat Hunting, February 2016. URL: https://www.sans.org/white-papers/who-what-where-when-why-how-effective-threat-hunting/.

[61] Robert Lemos. 5 Lessons Learned From OPM Data Breach, September 2016. URL: https://www.eweek.com/security/5-revelations-from-opm-data-breach-report/.

[62] M. Majid and K Ariffi. Success Factors for Cyber Security Operation Center (SOC) Establishment. In Proceedings of the Proceedings of the 1st International Conference on Informatics, Engineering, Science and Technology, INCITEST 2019, 18 July 2019, Bandung, Indonesia, Bandung, Indonesia, 2019. EAI. doi:10.4108/eai.18-7-2019.2287841.

[63] Aleksandr Matrosov, Eugene Rodionov, David Harley, and Juraj Malcho. Stuxnet under the microscope. ESET LLC (September 2010), 6, 2010.

[64] Sarah M. McClanahan. 169th Cyber Protection Team – Highly Capable, Always Ready, November 2019. URL: https://news.maryland.gov/ng/2019/11/06/169th-cyber-protection-team-highly-capable-always-ready/.

[65] Kathryn McIver. Closing the Revolving Door. Technical report, August 2022. URL: https://icitech.org/wp-content/uploads/2022/08/Closing-the-Revolving-Door.pdf.

[66] Md Nazmus Sakib Miazi, Mir Mehedi A. Pritom, Mohamed Shehab, Bill Chu, and Jinpeng Wei. The Design of Cyber Threat Hunting Games: A Case Study. In 2017 26th International Conference on Computer Communication and Networks (ICCCN), pages 1–6, July 2017. doi:10.1109/ICCCN.2017.8038527.

[67] Sadegh M. Milajerdi, Birhanu Eshete, Rigel Gjomemo, and V. N. Venkatakrishnan. POIROT: Aligning Attack Behavior with Kernel Audit Records for Cyber Threat Hunting. Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pages 1795–1812, November 2019. doi:10.1145/3319535.3363217.

[68] Demetrio Milea. Hypothesis in Threat Hunting, July 2017. URL: https://medium.com/@demetriom/hypothesis-in-threat-hunting-4bea5446e34c.

[69] Ayesha Naseer, Humza Naseer, Atif Ahmad, Sean B. Maynard, and Adil Masood Siddiqui. Real-time analytics, incident response process agility and enterprise cybersecurity performance: A contingent resource-based analysis. International Journal of Information Management, 59:102334, August 2021. doi:10.1016/j.ijinfomgt.2021.102334.

[70] Military Family Advisory Network. Effects of Moving on Military Families. URL: https://www.mfan.org/topic/moving-permanent-change-of-station/effects-of-moving-on-military-families/.

[71] Tim Nosco, Jared Ziegler, and Zechariah Clark. The Industrial Age of Hacking. In Proceedings of the 29th USENIX Security Symposium, August 2020.

[72] Megan Nyre-Yu, Robert S. Gutzwiller, and Barrett S. Caldwell. Observing Cyber Security Incident Response: Qualitative Themes From Field Research. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63(1):437–441, November 2019. doi:10.1177/1071181319631016.

[73] Megan M Nyre-Yu. Determining System Requirements for Human-Machine Integration in Cyber Security Incident Response. 10 2019. doi:10.25394/PGS.10014803.v1.

[74] Jon Oltsik. The Life and Times of Cybersecurity Professionals 2018. Research Report, 2019. URL: https://www.esg-global.com/hubfs/pdf/ESG-ISSA-Research-Report-Life-of-Cybersecurity-Professionals-Apr-2019.pdf.

[75] James Pettigrew and Julie Ryan. Making Successful Security Decisions: A Qualitative Evaluation. IEEE Security & Privacy, 10(1):60–68, January 2012. Conference Name: IEEE Security & Privacy. doi:10.1109/MSP.2011.128.

[76] Hussein Rasheed, Ali Hadi, and Mariam Khader. Threat Hunting Using GRR Rapid Response. In 2017 International Conference on New Trends in Computing Sciences (ICTCS), pages 155–160, October 2017. doi:10.1109/ICTCS.2017.22.

[77] Paul Ryan. Is apprenticeship better? a review of the economic evidence. Journal of Vocational Education & Training, 50(2):289–325, June 1998. doi:10.1080/13636829800200050.

[78] Desiree Sacher-Boldewin and Eireann Leverett. The Intelligent Process Lifecycle of Active Cyber Defenders. Digital Threats: Research and Practice, 3(3):1–17, September 2022. doi:10.1145/3499427.

[79] Sena Sahin, Suood Al Roomi, Tara Poteat, and Frank Li. Investigating the password policy practices of website administrators. In 2023 IEEE Symposium on Security and Privacy (SP), pages 552–569. IEEE, 2023.

[80] Johnny Saldaña. The coding manual for qualitative researchers. The coding manual for qualitative researchers, pages 1–440, 2021. Publisher: SAGE publications Ltd.

[81] Karen Scarfone. The Hunters Handbook. ENDGAME & Accenture. URL: https://cyber-edge.com/resources/the-hunters-handbook-endgames-guide-to-adversary-hunting/download/.

[82] Eric Schmid. Military families often have to move every few years. Critics say it's disruptive and unnecessary, April 2022. Section: Military. URL: https://wusfnews.wusf.usf.edu/military/2022-04-03/military-families-often-have-to-move-every-few-years-critics-say-its-disruptive-and-unnecessary.

[83] Cyborg Security. Cyber Threat Hunting - What Is It, Really?, May 2022. URL: https://www.cyborgsecurity.com/blog/cyber-threat-hunting-what-is-it-really/.

[84] Cyborg Security. The Threat Hunter's Hypothesis, March 2022. URL: https://www.cyborgsecurity.com/library/guides/the-threat-hunters-hypothesis-2/.

[85] Sally Smith, Ella Taylor-Smith, Khristin Fabian, Matthew Barr, Tessa Berg, David Cutting, James Paterson, Tiffany Young, and Mark Zarb. Computing degree apprenticeships: An opportunity to address gender imbalance in the IT sector? In 2020 IEEE

Frontiers in Education Conference (FIE), pages 1–8, October 2020. ISSN: 2377-634X. doi:10.1109/FIE44824.2020.9274144.

[86] Jonathan M. Spring and Phyllis Illari. Review of Human Decision-making during Computer Security Incident Analysis. Digital Threats: Research and Practice, 2(2):1–47, June 2021. URL: https://dl.acm.org/doi/10.1145/3427787, doi:10.1145/3427787.

[87] Geoff Stoker, Ulku Clark, Manoj Vanajakumari, and William Wetherill. Building a Cybersecurity Apprenticeship Program: Early-Stage Success and Some Lessons Learned. page 10, 2021.

[88] Sathya Chandran Sundaramurthy, Jacob Case, Tony Truong, Loai Zomlot, and Marcel Hoffmann. A Tale of Three Security Operation Centers. In Proceedings of the 2014 ACM Workshop on Security Information Workers - SIW '14, pages 43–50, Scottsdale, Arizona, USA, 2014. ACM Press. doi:10.1145/2663887.2663904.

[89] Sathya Chandran Sundaramurthy, John McHugh, Xinming Ou, Michael Wesch, Alexandru G. Bardas, and S. Raj Rajagopalan. Turning contradictions into innovations or: How we learned to stop whining and improve security operations. In Twelfth Symposium on Usable Privacy and Security (SOUPS 2016), pages 237–251, Denver, CO, June 2016. USENIX Association.

[90] Sathya Chandran Sundaramurthy, John McHugh, Xinming Simon Ou, S. Raj Rajagopalan, and Michael Wesch. An Anthropological Approach to Studying CSIRTs. IEEE Security Privacy, 12(5):52–60, September 2014. Conference Name: IEEE Security Privacy. doi:10.1109/MSP.2014.84.

[91] R Symonds. Innovating the Prioritization of Cyber Defense. Journal of Information Warfare, 16(2):12–18, 2017. Publisher: Peregrine Technical Solutions.

[92] Stoney Trent, Robert R. Hoffman, David Merritt, and Sarah Smith. Modelling the Cognitive Work of Cyber Protection Teams. The Cyber Defense Review, 4(1):125–136, 2019. Publisher: Army Cyber Institute.

[93] Rick Van der Kleij, Geert Kleinhuis, and Heather Young. Computer security incident response team effectiveness: A needs assessment. Frontiers in Psychology, 8, 2017. doi:10.3389/fpsyg.2017.02179.

[94] Rob van Os, Marcus Bakker, Ruben Bouman, Martijn Docters van Leeuwen, Marco van der Kraan, and Wesley Mentges. TaHiTI: a threat hunting methodology, December 2018. URL: https://www.betaalvereniging.nl/wp-content/uploads/TaHiTI-Threat-Hunting-Methodology-whitepaper.pdf.

[95] Kevin Wafula and Yong Wang. CARVE: A Scientific Method-Based Threat Hunting Hypothesis Development Model. In 2019 IEEE International Conference on Electro Information Technology (EIT), pages 1–6, May 2019. ISSN: 2154-0373. doi:10.1109/EIT.2019.8833792.

[96] Renzheng Wei, Lijun Cai, Aimin Yu, and Dan Meng. DeepHunter: A Graph Neural Network Based Approach for Robust Cyber Threat Hunting, April 2021. arXiv:2104.09806 [cs].

[97] Rodrigo Werlinger, Kasia Muldner, Kirstie Hawkey, and Konstantin Beznosov. Preparation, detection, and analysis: the diagnostic work of IT security incident response. Information Management & Computer Security, 18(1):26–42, March 2010. doi:10.1108/09685221011035241.

[98] Alan White and Ben Clark. BTFM: blue team field manual. Alan White, United States, version 1, rel 2 edition, 2017.

[99] David E Whitehead, Kevin Owens, Dennis Gammel, and Jess Smith. Ukraine cyber-induced power outage: Analysis and practical mitigation strategies. In 2017 70th Annual Conference for Protective Relay Engineers (CPRE), pages 1–8. IEEE, 2017.

[100] John R. Wilson and Sarah Sharples, editors. Evaluation of Human Work. CRC Press, Boca Raton, 4 edition, April 2015. doi:10.1201/b18362.

A Summary of the Technical Report

An extended version of this paper is available as a technical report at arXiv:2402.12252. It includes:

- The full interview protocol.
- Saturation charts showing that our $N = 11$ interviews saturated the coding space under consideration.
- Discussion of the possible effects of subjects' organizational ranks.
- The codebooks used in our analysis.
- Details about cybersecurity frameworks and related TH processes.
- Additional details of our methodology and results, notably a detailed TH process diagram refining the high-level version presented in Figure 2.