DISCOVERING U.S. GOVERNMENT THREAT HUNTING PROCESSES AND IMPROVEMENTS

by

William P. Maxam III

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Electrical and Computer Engineering

School of Electrical and Computer Engineering

West Lafayette, Indiana

May 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. James Davis

School of Electrical & Computer Engineering

Dr. Alexander Quinn

School of Electrical & Computer Engineering

Dr. Santiago Torres-Arias

School of Electrical & Computer Engineering

Approved by:

Dr. Dimitrios Peroulis

To my loving wife, Lauren. Without you, this thesis would not have come to be.

ACKNOWLEDGMENTS

I would like to thank my Lord Jesus Christ for the grace He has given me, my wife, Lauren, and son, Q, for their love, support and the joy they bring me, my father and mother, for their love and support in every stage of my life, Dr. Davis for mentoring me and teaching me everything I know about research and technical writing, Dr. Torres-Arias and Dr. Quinn for their feedback and guidance with this research, Geoff Cramer for dropping everything to provide a second review of my data, the rest of the Duality Lab for their constant support, feedback and fraternity during my time at Purdue.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

APT	Advanced Persistent Threats
ATT&CK	Adversarial Tactics, Techniques, and Common Knowledge
CG	Coast Guard
CGCYBER	Coast Guard Cyber Command
CISA	Cybersecurity and Infrastructure Security Agency
CPT	Cyber Protection Team
CTF	Capture The Flag
CTI	Cyber Threat Intelligence
DHS	Department of Homeland Security
DOD	Department of Defense
DOE	Department of Energy
DOS	Department of State
EDR	Endpoint Detection and Response
IDS	Intrusion Detection System
IR	Incident Response
IRB	Internal Review Board
ISAC	Information Sharing and Analysis Center
IT	Information Technology
NASA	National Aeronautics and Space Administration
NIST	National Institute of Standards and Technology
OPM	Office of Personnel Management
PAI	Publicly Available Information
RQ	Research Question

SIEM	Security Information and Event Management
SOC	Security Operation Center
TaHiTI	Targeted Hunting integrating Threat Intelligence [1]
TH	Threat Hunt/Hunting
TTP	Tactics, Techniques and Procedures
US	United States
VPN	Virtual Private Network

ABSTRACT

INTRODUCTION Cyber Threat Hunting (TH) is the activity of looking for potential compromises that other cyber defenses may have missed. These compromises cost organizations an estimated $10M each and an effective Threat Hunt can reduce this cost. TH is a new discipline and processes have not yet been standardized. Most TH teams operate with no defined process. This is a problem as repeatable processes are important for a mature TH team.

OBJECTIVES This thesis offers a Threat Hunt process as well as lessons learned derived from government TH practice.

METHODS To achieve this I conducted 12 interviews, 1 hour in length, with government threat hunters. The transcripts of these interviews were analyzed with process and thematic coding. The coding was validated with a second reviewer.

RESULTS I present a novel TH process depicting the process followed by government threat hunters. Common challenges and suggested solutions brought up by threat hunters were also enumerated and described. The most common problems were minimal automation and missing measures of TH expertise. Challenges with open questions were also identified. Open questions include: determining how to identify the best data to collect, how to create a specific but not rigid process and how to measure and compare the effectiveness of TH processes. Finally, subjects also provided features that indicate expertise to TH team members and recommendations on how to best integrate newer members into a TH team.

CONCLUSION This thesis offers a first look at government TH processes. In the short term, the process recommendations provided in this thesis can be implemented and tested. In the long term, experiments in this sensitive context remain an open challenge.

1. INTRODUCTION

Computer network security is a challenge in the modern world. Cyber intrusions are a concern both for governments and private corporations. Unauthorized network infiltrations cost single organizations an average of $13 million a year [2]. Governments have additional non-monetary concerns, such as protecting election systems and maintaining national security [3].

The longer an adversary dwells undetected on a network, the more damage the adversary can cause. An Aberdeen analysis found that a 50% reduction in dwell time would reduce the cost of an attack by ∼30%[4]. IBM found that data breaches cost on average $1.12 million more if not contained within the first 200 days [5]. These costs estimations include lost revenue, regulatory and legal fees, and the technical cost of forensics activities [5]. They estimated the average adversary dwell time at 230 days, not including an additional time to respond to the breach [5].

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*" [6]. The government routinely performs TH on contractor's networks [3] and the private sector is establishing this capability [7]. In a 2017 survey across Telecommunications, Tech, Government, Healthcare and Financial institutions, most organizations engaged in threat hunting [8], but as a discipline TH is still in its infancy.[1] However, the survey showed that less than half of organizations that perform TH use a written TH process [8]. The processes currently used by TH teams are not well documented [1] and the current state of research in hunt teams does not describe their processes in detail [9]–[11]. This limits the team's ability to adopt best practices and improve their processes over time and it also limits how well academia can assist TH teams.

In addition to a lack of process documentation, most cyber security teams, TH teams included, experience a high amount of turnover [12]–[14]. The median tenure is ∼ 2 years [15].

[1]↑The publisher of this survey, SANS, is a premier cyber security training organization so the sample likely contains primarily large/government organizations. Company size is not discussed. The people surveyed were also IT or cyber security employees so the sample is more likely to include information from organizations with large cyber security departments.

Providing process documentation is generally important for new members [16] and even more so if onboarding is done in a remote setting [17]. Processes have also been shown to use new members more efficiently in other cyber security tasks [18].

My thesis is a starting point to address these concerns. This thesis presents an example Government TH process. In addition, this thesis provides TH recommendations and lessons learned derived from US government Threat Hunt teams. These results can be used as a springboard to study many unknowns within threat hunting.

My approach was to interview 12 threat hunt practitioners across three different United States government organizations. The interview protocol I followed had 2 focuses: (1) the TH process and the challenges they experience with the process, and (2) the way new members interact with the process. Thematic coding was used to create a representative cyber threat hunt process and understand the common concerns of the subjects.

These results are presented to be usable for practitioners and researchers. I introduce the first published Threat Hunting process model for Government TH teams (Figure 4.1). The thematic analysis identified common challenges, their recommended solutions, and open questions in TH processes (Table 4.2). Many open questions remain in the TH landscape and are discussed along with unexpected findings in chapter 5.

2. BACKGROUND

2.1 Threat Hunting

2.1.1 Threat Hunting in the Cybersecurity Landscape

Multiple proactive cyber security teams exist including Blue Teams, Compliance Teams, and Security Operations Centers (SOCs). Blue teams seek to protect the network from intruders by hardening the network's security [19]. Compliance Teams enforce cybersecurity best practices across a network [20].

SOCs may have some overlap with the two previously mentioned teams but are primarily responsible for detecting adversaries [21]. SOCs have received much attention from researchers. Researchers have both interviewed SOC members [22] and been embedded in SOCs [23]. Classification systems exist for SOCs [24], best practices have been enumerated [25] and SOC processes are being studied [26].

Incident Response (IR) teams, when separate from SOCs, are a reactive cyber security team responsible for evicting adversaries that have been found on the network [27].

Sometimes IR teams react to adversaries regardless of how they were detected. A SOC may detect an adversary or a company may receive a notification that they have been compromised [28], but either way the IR team will be called to remove the adversary. IR teams also receive much attention from academia and government agencies [29] Researchers have embedded themselves in IR teams [27], [30] and have interviewed IR team members [31], [32].

Proactive offensive security teams include Red Teams and Penetration Testing Teams. These teams purposefully take offensive actions on the network in order to simulate adversaries. Penetration Testing Teams typically look only for vulnerabilities at the network edge [33]. Red teams seek to infiltrate past the network edge and further into the network [33]. Red teams often operate in conjunction with a Blue Team [33].

Threat hunting (TH) teams have a unique role compared to these cybersecurity teams. Unlike red teams or penetration testers, they perform no offensive actions [34]. Unlike blue teams or security compliance teams, they do not harden the network [34]. They only search the network for adversaries that an organization's previously existing defenses may have

missed [35]. They hunt on the assumption that an adversary has infiltrated the network when there is no sign of a network compromise [11].

If an adversary is ever able to compromise a network undetected, it could cause significant damage [4]. In situations when adversaries are not immediately detected, intrusions become more expensive [4], [5]. To avoid undetected adversaries on the network, TH teams search for any adversaries that have evaded detection by usual methods [6]. This is analogous to a military unit that does not only rely on its gate guards (the SOC) but also has units patrol inside the camp hunting for adversaries that may have circumvented the usual security protocols at the gate.

Figure 2.1 depicts such a situation where an adversary has infiltrated the network past the proactive defenses erected by the cooperation of Red, Blue, Penetration Testing and Compliance Teams. The adversary is not within the network and avoided detection by the SOC. Three examples of how such an adversary may be detected after already evading the security at the network edge are shown in Figure 2.1. The top path depicts a TH team, hunting inside the network finding the adversary. However, this is not the only way to detect compromises that have initially gone undetected. The adversary may be caught by one of the SOCs internal network or host based sensors (middle path) or an agency like the FBI may notify the organization that they have been compromised (bottom path). It does not matter how the adversary was detected, once a detection has occurred, the IR team is called to evict the adversary.

Although TH is a new field [1], [36], it is considered an important cybersecurity capability [37], [38]. TH is mandated for Federal Civilian Executive Branch Agencies [39] like the National Aeronautics and Space Administration (NASA), Department of State (DOS) and the Department of the Treasury [40]. Since TH is an emerging field of cyber security, it has not yet been widely researched. As such, I make use of gray literature to collect information for this background section.

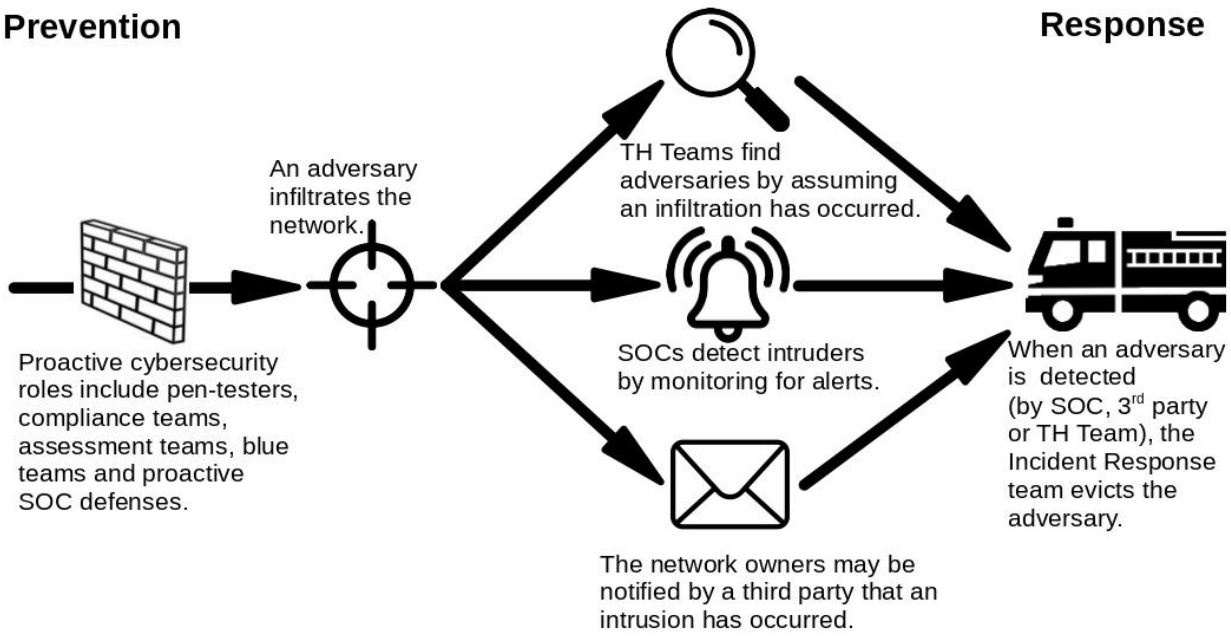


Figure 2.1. Three paths exist to discover an adversary.

2.1.2 Threat Hunting 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, organizations either designate SOC personnel to perform hunting or maintain teams dedicated to hunting [6]. External organizations that offer THs on private sector networks include private sector organizations such as Booz Allen Hamilton [41], Crowdstrike [42] and Cisco [43]. Most organizations opt for internal hunt teams [8].

2.1.3 Threat Hunting in the Public Sector

In the United States, the government uses both civilian and military teams. For example, the Cybersecurity and Infrastructure Security Agency's (CISA) Hunt and Incident Response Team (HIRT) [44] is a civilian team that does TH. Cyber Protection Teams (CPTs) are military teams that exist within all branches of the military [45] as well as many states' National Guards [46]. Some of these CPTs do TH along with other cybersecurity functions [47]. Many

government teams are deployed to federal civilian networks [39], military networks [48], and even private networks [3], [44].

2.1.4 Common Problems Affecting Threat Hunt Teams

Both government and Private Sector Threat Hunting teams deal with a high turnover of personnel. All cyber security organizations, public and private, deal with half of cyber security professionals not staying in a job longer than two years [15]. In addition to this turnover, military organizations often rotate active duty personnel every two to three years [49]–[51]. Some military branches cannot fill the cybersecurity positions they have available [52]. Good cybersecurity processes have been shown to mitigate the adverse effects of less expert personnel [18], helping mitigate any adverse effects personnel turnover may create.

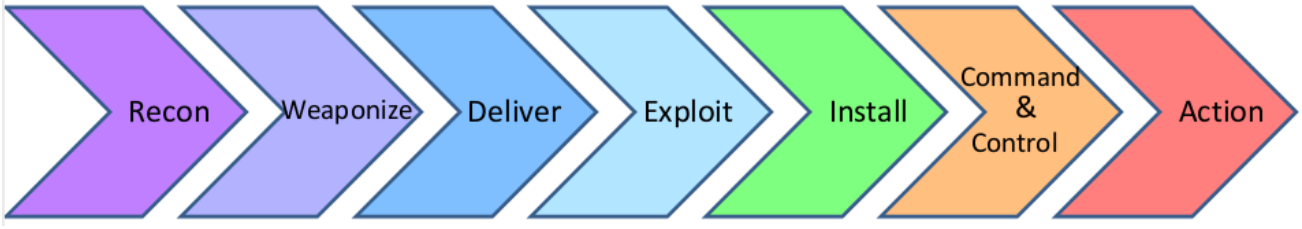
2.2 Threat Hunting Frameworks, Models, and Processes

2.2.1 TH Frameworks

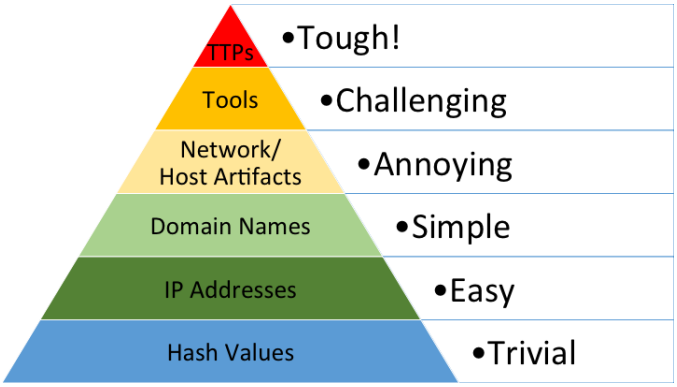
Currently, the most common way to hunt for an adversary is ad hoc, *i.e.,* without a formal process [8]. This is true despite the fact that multiple frameworks have sought to provide structure for TH teams. Both academia [11] and private sector [1] documents indicate the use of three popular frameworks: Lockheed's Kill Chain [53], the Mitre ATT&CK framework [54], and the Pyramid of Pain [55]. 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 (see Figure 2.2a) and ATT&CK frameworks outline the steps an attacker takes to carry out a successful attack; ATT&CK goes into greater detail.[1] The Pyramid of Pain (see Figure 2.2b) is organized differently and instead assesses what information is most valuable for disrupting adversary activity.

Academic researchers often suppose a framework called the *hypothesis method* [9]–[11]. 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. For example

[1]↑The Mitre ATT&CK Framework can be viewed at https://attack.mitre.org.



(a) Lockheed's Cyber Kill Chain [56].



(b) Bianco's Pyramid of Pain [57].

Figure 2.2. Two example Cybersecurity Frameworks. TTPs are Tactic, Technique, and Procedures

a hypothesis may be: APT #X used a known exploit to compromise a VPN Server then moved laterally to exfiltrate information from a certain database.

All the frameworks and methods in this section can be used in tandem with each other [1]. 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.

2.2.2 Private Sector Threat Hunt Processes

Four organizations from the Dutch financial sector shared a process known as the Targeted Hunting integrating Threat Intelligence (TaHiTI) methodology [1]. TaHiTI was created because van Os *et al.* noticed that the TH community lacked standardized processes and definitions. TaHiTI does not attempt to describe government hunt teams. It provides a TH process derived from a private sector round table, describing what should occur, not

necessarily what does occur. TaHiTI's process is depicted in Figure 2.3. Typically, academic works on TH methodologies do not seek to describe current TH practices and instead focus on novel approaches that they believe could assist TH teams [9], [10]. It is also important to note TaHiTI does not seek to provide a process for data-driven only structured, hypothesis-based TH.

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* [58] with an accompanying handbook for managers [59], but these were not as detailed as TaHiTI. Other TH guides did not even bother attempting to provide a TH process [60].

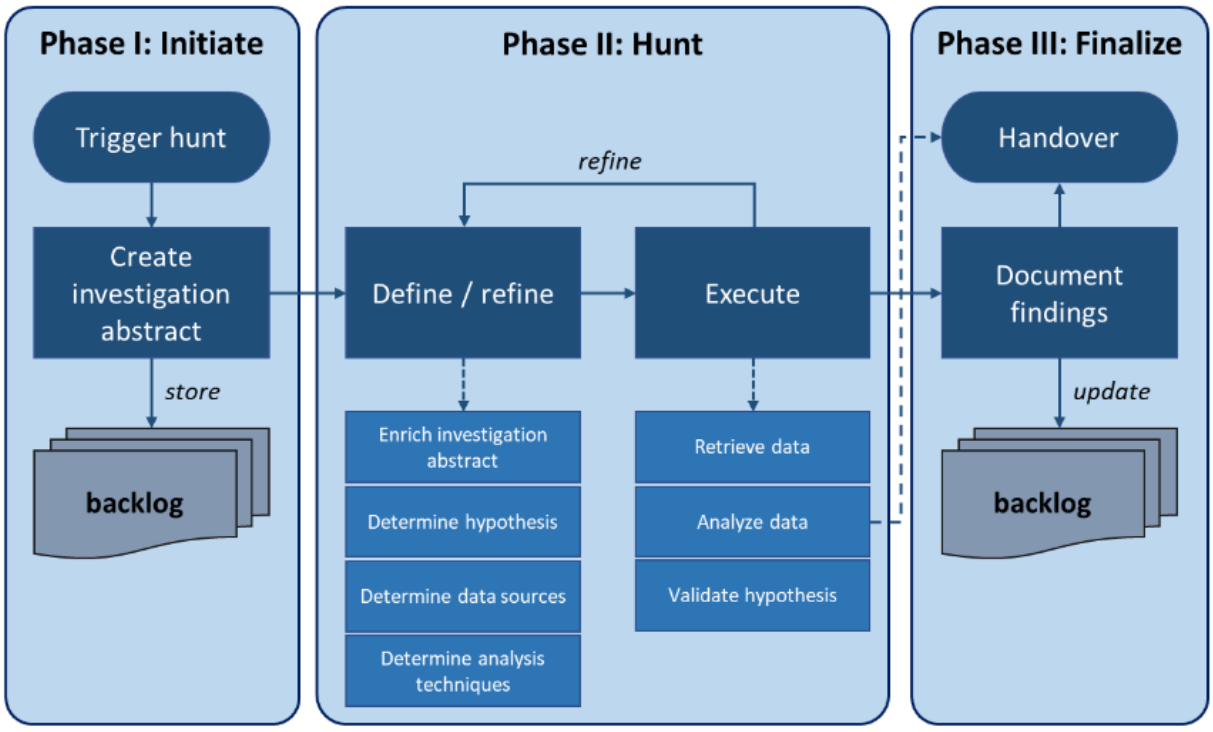


Figure 2.3. The TaHiTI process from van Os *et al.* [1].

2.2.3 Public Sector Threat Hunt Processes

Currently, little is known about government TH processes. While the private sector has attempted to share hunt methodologies [1], the government process has not been as accessible. To my knowledge, there is not attempt, focused only on TH, to document the

process used by government TH teams. The fact that the government is among the sectors most often targeted [61] makes the government's defenses, including TH, of increased interest.

The closest document to a government TH process was work done by Trent *et al.* [62]. This study created a *cognitive* model of activities performed by US Army CPTs. Their goal was not to outline a Threat Hunting *process* but CPTs do perform TH so the findings are relevant to an attempt to report an example TH process. The cognitive model found by Trent *et al.* can be seen in Figure 2.4.

In the past government cybersecurity practice has influenced private organizations. Historically, the military has driven the creation of informal cybersecurity standards [63]. Additionally, after the 2015 Office of Personnel Management (OPM) data breach [64], many private organizations sought to learn from the government [65]–[68]. In this case an example government TH process may be beneficial for the same reasons.

2.3 Comparing Prior Works

There exists substantial differences between Trent *et al.* [62] and TaHiTI [1]. The comparison is not perfect as TaHiTI was focused only on TH but Trent *et al.* covered all three of the CPTs' mission sets, only one of which is similar to a TH. Additionally TaHiTI is a process while Trent *et al.* is a cognitive model. However, these works indicate that differences exist between the private sector processes and work done by military Cyber Protection Teams (CPTs). For example, TaHiTI is explicitly a structured hunting process based on the hypothesis method and does not focus on unstructured data-driven hunting. The difference is that structured hunting has a hypothesis that can be disproven, whereas unstructured hunting relies on analysts looking through data without a hypothesis. Trent *et al.* gives no explicit indication of a hypothesis being used, even though generating a hypothesis would be a unique cognitive task.

2.4 Measuring Threat Hunt team Effectiveness

There exist a few metrics that can be used to measure TH team effectiveness. The inventor of the Pyramid of Pain, Bianco, also provided a Hunt Maturity Model [69]. The

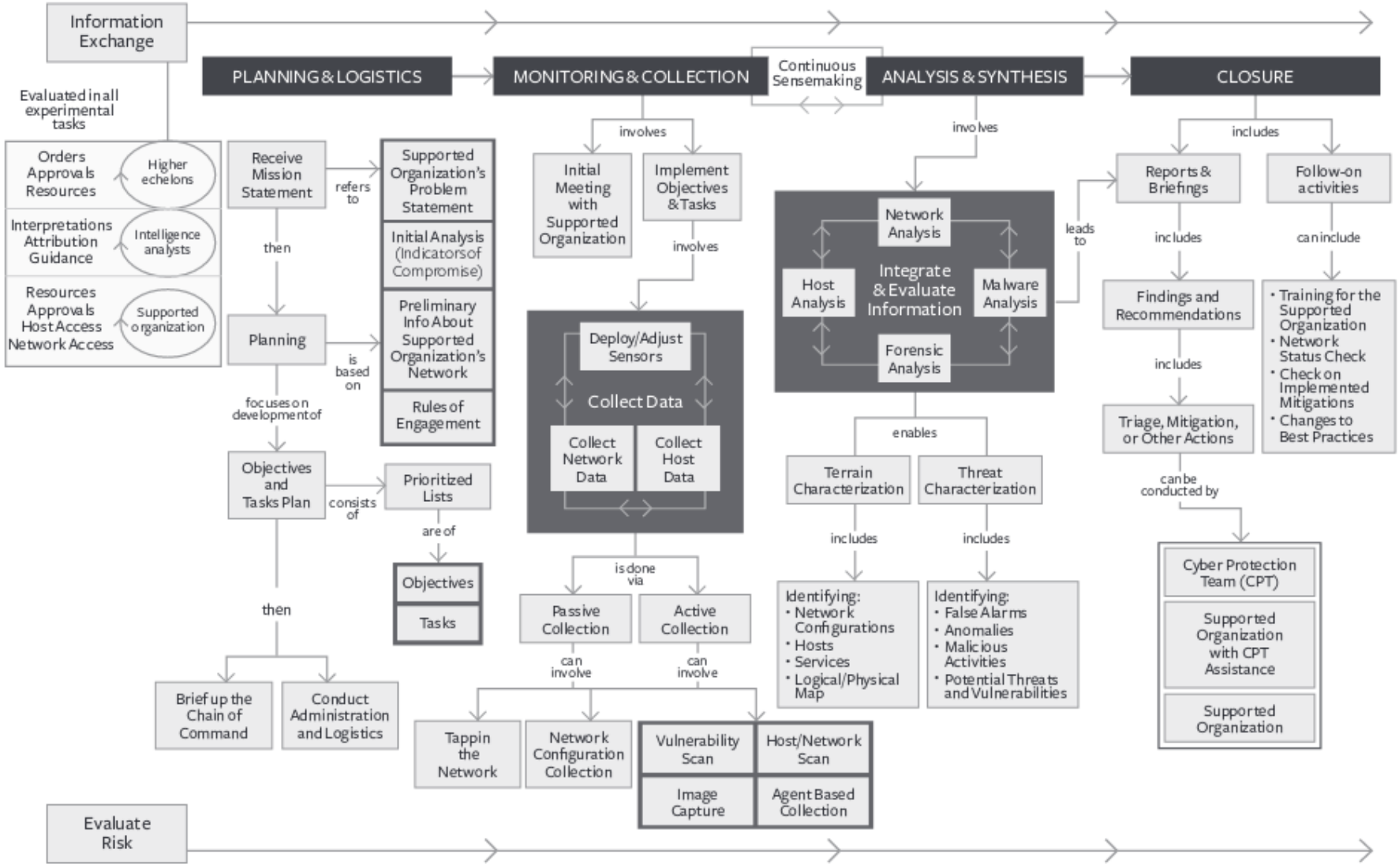


Figure 2.4. Detailed CPT cognitive work model from Trent *et al.* [62]

model describes the differences between 5 different levels of TH teams ranging from level HMM-0 (Initial) to level HMM-4 (Leading). Bianco's level consider factors like process reproducibility and how intelligence is used. Levels are depicted in Figure 2.5. TH teams at HMM-0 are not considered to be capable of hunting. Other recommendations for measuring a TH effectiveness include metrics like dwell time [1] or how much the TH team improved security [58]. More generally risk reduction has been proposed as a method for measuring the effectiveness of cyber security practices [70]. There is no agreement on the best metrics for measuring TH team's effectiveness.

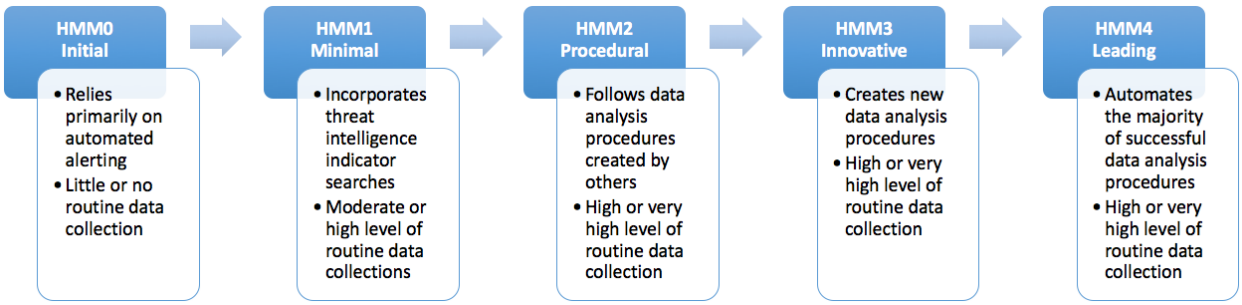


Figure 2.5. Bianco's Threat Hunting Maturity Model [69].

2.5 Threat Hunting Tools

Two surveys conducted by Rob Lee & Robert Lee [8] and Robert Lee & Rob Lee [71] provide insight into the tools used by TH teams. Almost all teams consider endpoint security data and access logs critical to a Threat Hunt with over 90% of teams using existing infrastructure tools such as a Security Information and Event Management (SIEM), Endpoint Detection and Response (EDR), or Intrusion Detection System (IDS). About half of the teams use custom tools while less than half of teams use open-source TH software. Although technology is the highest cost budget item for most TH teams, less than a third use third-party TH tools. The reason why so few TH teams invest in TH-specific tools is not clear in either survey.

Despite more TH teams relying on tools that are not TH-specific, in recent years many TH tools seeking to automate the TH process have been created in both the private sector

and academia. Researchers have proposed different tools to support different TH functions. Some tools focus primarily on the detection of adversaries using novel techniques. For example, Poirot uses graphs to model and match relationships between indicators of compromise and relationships between system components [72], and DeepHunter uses a neural network to do similar graph pattern recognition [73]. Other tools focus on workflow automation. For example, TH teams use their knowledge of intelligence reports to search for adversarial activity in system logs. This conversion of human-readable reports to specific search queries has been a target for automation [38], [74]. Lastly, some academic works propose adapting tools traditionally created for Incident Response teams [36] or traditional network defenders [75] as TH tools.

Similarly, often private sector tools marketed as TH tools are geared primarily towards SOCs [76], [77] as opposed to standalone TH teams. As mentioned previously, the reason why so few TH teams invest in TH-specific tools is not clear. It could be a cost-saving mechanism or because SOC tools are already readily available on most networks. It may be that proposed TH tools or automation frameworks assume certain models of threat hunting [78], [79] which may or may not accurately reflect current TH practice. Additionally, some TH solutions blur the line between a SOC performing continual monitoring and the goals of a TH team, making use of the tool unrealistic for a third-party TH team [75]. For example, a tool may automatically perform a mitigating action that a SOC analyst normally would perform while the TH team may not authorized to take those actions without approval from the network owner.

2.6 Summary and Unknowns

TH teams perform a unique function, hunting for adversaries internal to the network boundary. TH teams are relatively new and have not yet been given much attention by researchers. Currently, most TH is done without a process [8]. Very little is known about TH team's processes in general. The most detailed process was provided by TaHiTI, a process shared by threat hunters from four dutch financial institutions [1]. While prior work provides a private sector TH process and a cognitive model of CPT missions, the process

used by government TH teams is unknown. Measuring TH team effectiveness is still an open questions and TH teams seem to use open-source, custom, or pre-existing tools more than investing in TH specific tools.

3. METHODOLOGY

3.1 Research Questions

My goal is to describe the Threat Hunting (TH) process used by government TH teams and how that process interacts with less expert members. The goal of this thesis is not to prescribe a process, as measuring the effectiveness of TH processes is an open question (see §2.4). Assessing the quality of the described process is left to future work.

As discussed, most organizations that perform TH, do it *ad hoc* . Outside of TH, processes have been used to better integrate less expert cyber security professionals [18] and thereby reduce the cost of having a rotating workforce. A published example government TH process would provide a sample process that some of these organizations may consider as a starting point. Organizations that have a process could compare their own to this sample process in an effort to recommend improvements, either to their own processes or published to help other organizations. Any observed lessons learned and challenges observed by this study can be similarly used as a starting point for organizations without a process or for comparison against TH processes used by organization with processes. Where no solutions exist for challenges government TH teams are having, academia can use this study as a launching point for future research and to propose improvements. With these results, researchers may also be able to focus future TH research on areas where government teams are experiencing challenges. Since our study is focused on government TH teams — teams that often have high turnover rates (see §2.1.4) — our research is also interested in documenting how the government TH process integrates less expert members and what determines expertise in the TH domain. These issues could be similarly helpful to researchers and practitioners. To this end, I studied two research themes each further subdivided into 2-3 research questions:

Theme #1: Current government TH processes

- **RQ1.1:** What processes are currently used by government TH teams?
- **RQ1.2:** What shortcomings exist with current government TH processes and what can be done to alleviate these shortcomings?

Theme 1 is focused on the TH processes currently implemented by Government TH teams. By discovering this process, further TH research can be conducted and other TH organizations can use the government process as a starting point or a second process to compare theirs against. Answers to RQ2.2 are important because these shortcomings could also exist in other TH processes, so other organizations may be able to learn from what Government teams are doing to alleviate the issue. If no solution to a given shortcoming exists, those issues would be good directions for future research.

Theme #2: New member integration

- **RQ2.1:** How do newer members fit into government TH processes?
- **RQ2.2:** How could government TH process changes facilitate the integration of less expert members?
- **RQ2.3:** What features indicate expertise to government TH team members?

Theme 2 is focused on the aspect of turnover within cyber security organizations and specifically government TH teams. Presumably, government TH teams are, somehow, currently integrating new members. How they are doing this could be of interest to other teams and, if they are having difficulties, could become directions for future research. RQ2.2 is included to make specific recommendations to government TH teams, although they also may inform other TH teams. RQ2.3 is included to define what expertise in TH looks like. Answers to this question could help government TH teams identify experts not only internally but also externally for recruiting and hiring. Answers may also inform future work to determine if what government TH team members perceive as expertise affects the team's success.

3.2 Design

Little information exists on the processes used by government TH teams so an exploratory study is needed to direct future research. The Semi-Structured Interview Study methodology was chosen for this study as a result of the population being studied. Such an exploratory study would be difficult to conduct using any other methodology. It would be difficult to

capture the complexity of the process using a survey study and since the population of government threat hunters is small and difficult to access, a meaningful sample size would be difficult to achieve. A survey would also not allow for iterative improvement on the collection instrument to pick up unexpected intricacies. A grounded theory methodology can introduce a new theory where no prior theory or framework exists, but given the prior work done by Trent *et al.* [62] and van Os *et al.* [1], an interview study was chosen because I felt these works provided enough initial structure to frame the collection instrument and initial analysis.

3.3 Recruiting

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. Participants were recruited for interviews by reaching out to the researcher's professional network. During recruitment, emails were sent to TH members with varying levels of experience and various positions. The response rate from direct contacts was 59% (10/17), and two additional subjects reached out to the research team, offering to be interviewed after hearing about the study. All participants were volunteers and not compensated. 11 out of the 12 participants had worked with me previously as peers (4), subordinates (5), or managers (2). My previous role on a TH team allowed me access to these TH team members and helped ensure no sensitive information was disclosed.

3.4 Ethics and National Security

This study was approved by Purdue University's Internal Review Board (IRB). A signed consent form was collected from each participant before their interview. All three organizations studied are small communities so special care was taken that none of the subjects could be identified. For example, in one organization I studied, there only exist two TH team members acting in a certain capacity and less than five members with as much TH-specific experience as the most experienced subject in my sample. For this reason, job descriptions are not connected to any subjects specifically. Similarly, quotes were not tied to a specific

subject. Instead, demographic information such as job role, is provided in relation to the quote without de-anonymizing any subjects.

Due to the sensitive information being discussed, I took care to ensure no classified information or customer information 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. Each transcript was then reviewed again for sensitive information, by the research team before being used for analysis.

3.5 Subjects

Table 3.1 shows each subject's TH experience (in years) and the number of TH missions each subject has been engaged in. As precise titles could de-anonymize subjects, subjects are mapped to three groups: Leadership, Team Leads, and Analysts. *Analysts* are TH team members that do technical work during Threat Hunts. *Team Leads* are intermediaries that manage analysts, coordinate analysis, and report to members of leadership. Not every organization had members that fit into this category. *Leadership* are hunt team leaders who do not perform analysis specific to a single Threat Hunt but instead manage TH teams and team leads or work on improving processes, logistics, analytics or tools. This category includes analyst's managers, officers in command, and executive, operations, and analysis officers. Subjects broken down by these positions can be seen in Table 3.2.

These three categories also generalize to private sector teams. Since the Analysts role exists in the private sector [1], the Leadership role would also exist and refer to their managers. Larger organizations may have a multi-tiered TH team structure and thus they would also have the Team Lead group although it is likely less universal than the other two groups. This is also what I observed in government teams.

The breakdown of subjects by their organization can be seen in Table 3.3. Some subjects had operated on teams in multiple organizations within the last three years so the numbers in Table 3.3 do not add up to 12.

Table 3.1. Subjects' experience and missions.

Subject	Years of Experience	Number of Missions
1	5	10
2	2	6
3	6	24
4	< 1	1
5	< 1	5
6	1	3
7	1	0
8	4	2
9	7	5
10	1	1
11	< 1	1
12	2	2

Table 3.2. Subject breakdown by position.

Position	# subjects	Position Definition
Leadership	4	Managers of TH teams or senior members that work on improving processes, logistics, analytics or tools
Team leads	4	Intermediaries that manage analysts, coordinate analysis, and report to leadership
Analysts	4	TH team members that do technical work during a hunt

Table 3.3. Subject breakdown by organization. DHS: US Department of Homeland Security. DOE: US Department of Energy.

Organizations	# subjects
Coast Guard Cyber Command (CGCYBER) (DHS)	10
CISA (DHS)	3
Sandia National Labs (DOE)	1
Private Sector	1

3.6 Interview Procedure

The initial interview protocol was created using the related TH literature and informed by my two years of Threat Hunting experience working for the Department of Homeland Security. The main body of the instrument was created with two concerns in mind: (1) Previous TH literature and (2) Common process concerns. Previous TH literature indicated that I should ask questions about automation and analysis frameworks. The differences between TaHiTI and Trent *et al.*'s model indicated that questions related to process detail should be added, as TaHiTI is a very high-level process and Trent *et al.* model is much more thorough. Common process concerns included questions about new member integration, process creation and process modification.

The instrument had four sections (chapter A). First, I reminded subjects not to discuss sensitive information. Second, I collected demographic information. The third section focused on TH processes, asking subjects to provide a process flow diagram and then use it to describe their team's process. The diagram was often referenced throughout section three. The fourth section examined how new members and expert members fit into the process.

Following the recommendations made in [80], two internal pilot studies were conducted with the primary researcher being interviewed by other members of the research team. Following these internal pilots, 5 questions were added (2 questions and 3 follow-up questions) and one follow-up question was re-worded. Best practices for pilot studies were also observed during the first two interviews, however, since there was little change in the instrument, these pilot interviews were counted toward the rest of the study. Over 92% of the questions (22/24) were held constant over the entire study. 2 follow-up questions were removed. 2 questions and 9 follow-up questions were added to the protocol (marked with an asterisk in chapter A). These additions occurred when the research team observed phenomena that were not covered by the protocol. For example, early participants referenced a document called a "Mission Plan" and a question was added asking about its utility. Questions were added only if the topics emerged early in the interviews.

Each interview lasted $\sim$ 1 hour and was conducted over Microsoft Teams. Both audio and video were recorded, however, only the audio and the process diagram created by the

participants were used in the analysis. Prior to every interview, a signed consent form was signed by the participant.

3.7 Data Analysis

Two types of analysis were performed: (1) process coding was used to answer RQ1 and (2) thematic coding was used for both RQ1 and RQ2.

Process coding was done as described in Saldana [81]. First, both TaHiTI and Trent *et al.*'s model were used to create separate codebooks. TaHiTI's codebook was generated using Figure 2.3 and the list of triggers provided in TaHiTI. Trent *et al.*'s codebook was generated using Figure 2.4. The subject descriptions of the process were then coded against each codebook.

After process coding was complete, the entire transcript was then re-coded using thematic coding as described in Guest *et al.* [82]. Memos were written by me and attached to the raw transcripts. These memos (636) were arranged into 58 themes and each theme became one of the top-level codes in the codebook. These codes were sometimes later divided into sub-codes to assist with analysis, *e.g.,* the 'Automation' code had multiple sub-codes, one of which was 'things that hinder automation'. All coding was done primarily by me with 6 unbroken portions from half the transcripts being provided to a second reviewer to measure inter-rater agreement. These 6 transcript portion were chosen to provide full coverage of the interview questions. They contained 56 coded sections each of which was blindly re-coded by the second reviewer. The result was a Kappa value of 0.82 .

Saturation was measured after all 12 interviews were complete. We measured saturation the same way it was measured in Guest *et al.* [83], by looking at the number of new codes appearing in each interview. The number of cumulative codes observed in each interview and in the previous interviews, charted in Figure 3.1, indicates that saturation was reached, after seven subjects with no new codes being observed in the last five subjects. All three organizations and all three position categories had been represented at that point, likely indicating a large amount of homogeneity within organizations. Codes on a per-interview

basis were also charted in Figure 3.2, showing that each interview provided at least 29 codes even if none were unique.

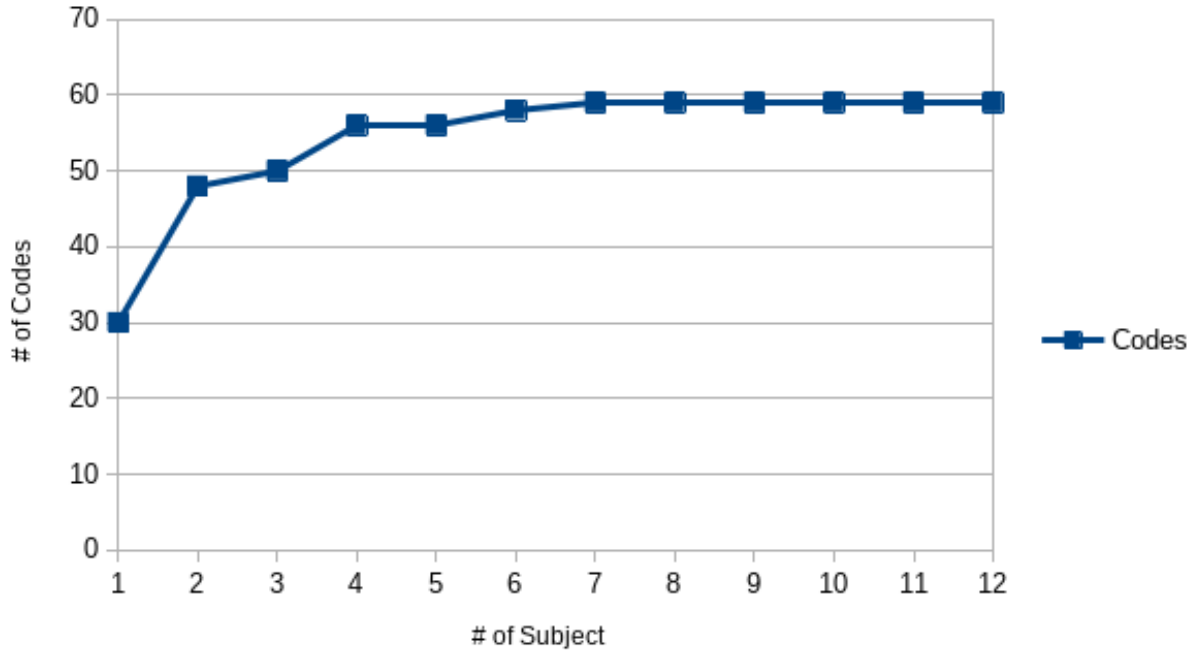


Figure 3.1. Cumulative unique codes by subject. The graph indicates saturation was achieved after 7 subjects.

3.8 Limitations and Threats to Validity

Similar to many other qualitative studies, the primary limiting factor in this thesis' generalizability is the sample size, both in terms of the number of subjects (12) and especially the DOE sample (1). However samples of this size (12) are not uncommon [84]–[86]. Guest *et al.* show that sample sizes as low as 6 could be sufficient if the population is homogeneous and the data being collected is specific [83]. I believe this is the case for this thesis. However, the process and results likely reflect the process of the CGCYBER most closely since most subjects were from that organization. The limiting factor of a small sample size was also mitigated by the diverse sample in terms of experience and TH team role.

Each of the three organizations studied had small populations. Therefore, the total pool of potential subjects was already limited. There are 39 people on a military Cyber Protection

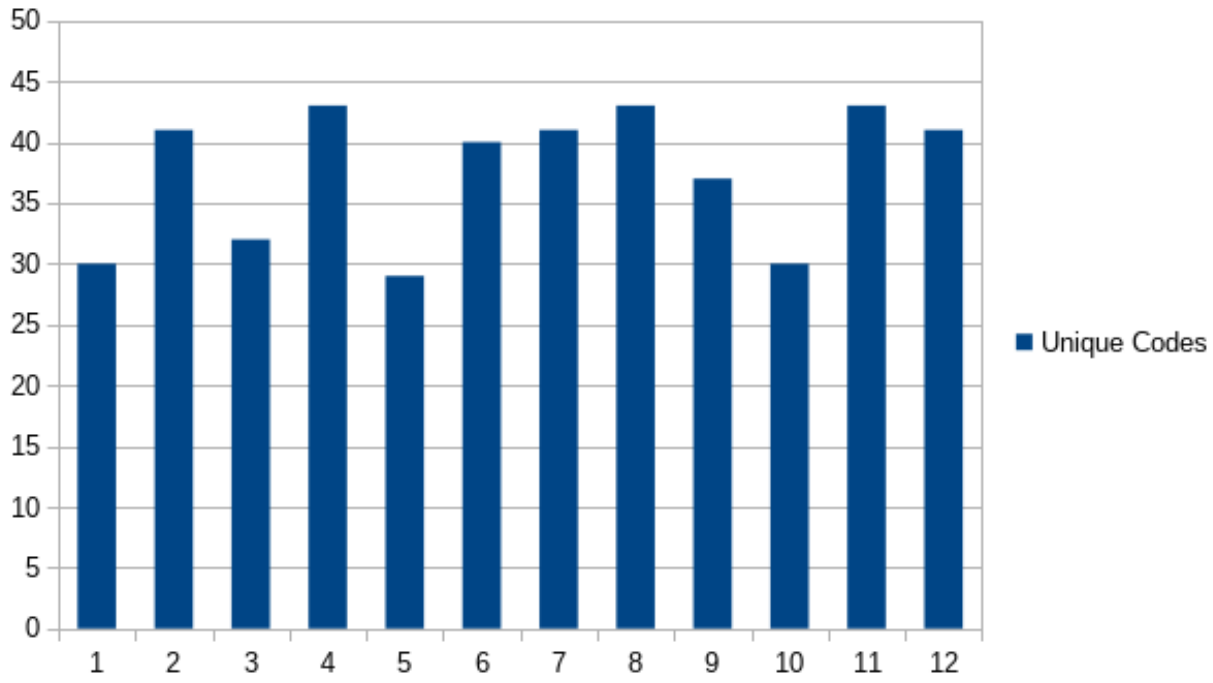


Figure 3.2. Unique codes observed per subject. The graph indicates that re-ordering the saturation chart (Figure 3.1) would not affect the saturation curve.

Team [87] and at the time of data collection, the Coast Guard had two teams [88]. The size of CISA's pool of Threat Hunters is not public but CISA is an agency with roughly 2,500 [89] employees only a small fraction of which are on an operational TH team.

The results may not generalize to the rest of the government or to private sector teams. Although the Coast Guard uses the same qualification Cyber Protection Team (CPT) guidelines, the operational differences in the teams may limit the study's generalizability to the CPTs of other branches and national guard CPTs. Every branch has a different use case, infrastructure, and culture. For example, the Coast Guard is a military branch under the DHS so they may have a closer relationship with CISA, another DHS agency, than the US Navy, which is under the Department of Defense (DoD). Since this thesis is focused on government hunt teams, findings may not be generalizable to teams outside the public sector.

A second limiting factor was the fact that the analysis was done primarily by one researcher. This was mitigated by strictly abiding by coding definitions. Inter-rater agreement

was measured using a second researcher matching 56 sections of coded text across 6 transcripts to the 58 codes in the codebook, resulting in a Kappa value of 0.82 .

Lastly a qualitative study relies on self-reporting. No study can fully eliminate bias and all subjects may have organizational or vocational blind spots. The results describe what subjects report and the report may not necessarily correspond perfectly to ground truth. Although this is the case, I was careful to follow best practices in interview instrument creation and the methodology followed is often used for exploratory studies [80]. I circulated my results to two experienced subjects (one team lead and one member of leadership). They both felt properly anonymized and represented by the work.

4. RESULTS

4.1 RQ1.1: What processes are currently used by Threat Hunt teams?

Each subject provided two types of data to indicate the process they used. First, while being interviewed 11 out of 12 subjects provided a process diagram at the beginning of their interview. Second, they described the process that their team used during the remainder of the interview.[1]

Initially, these verbal process descriptions were coded to the processes described by TaHiTI [1] (Figure 2.3) and Trent *et al.* [62] (Figure 2.4). However, both processes proved sufficiently different from the process being described by the subjects that neither could be used to code the data in a representative way. A synthesis of the observed processes is discussed in §4.1.3.

4.1.1 Coding to TaHiTI

I started by coding the interviews to the TaHiTI process because TaHiTI is the most detailed TH process in TH literature. This was done to see how closely the government TH process matched the process recommended by private sector TH practitioners. TaHiTI's methodology was not a good match to the process described by the subjects for two reasons: 1. TaHiTI's methodology assumed that the hunt team would be an internal team. 2. TaHiTI's methodology explicitly only describes structured, hypothesis-based threat hunting.

First, some tasks that are only required for external organizations were not included in the TaHiTI methodology. For example, 10 subjects mentioned placing sensors onsite prior to the team arriving on site.[2] TaHiTI does discuss data sources and ensuring data is available to the team but its concerns are different than my subject's concerns about sensor placement. Sensor placement or "deployment" is never discussed by TaHiTI but important to the subjects because both CGCYBER and CISA hunt teams operate primarily on external networks, not owned by their organization.

[1]↑The remaining subject did not feel sufficiently well versed with the process to draw a process diagram.
[2]↑subjects often use the term "*deployment*" to refer to the sensor placement or the team's arrival to the location of a hunt.

Second, 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 Threat Hunting team, being deployed on an unfamiliar network, takes time to document what processes are normal and expected. 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. Baselining was also not discussed by the TaHiTI process.

4.1.2 Coding to Trent *et al.*

Despite the fact the Trent *et al.* was a cognitive model for a CPT, not a process and not focused on TH, I coded the interview to the Trent *et al.* model to see how closely the model fit the process described by TH practitioners. Trent *et al.*'s cognitive workflow model was not observed to be a good match to the process described by subjects for two reasons. 1. Trent *et al.*'s work model was designed to distinguish between cognitive tasks. 2. Trent *et al.*'s work model describes all three of a CPT's missions not just TH missions.

First, subjects often draw administrative or temporal distinctions between tasks while Trent *et al.*'s distinguished between cognitive tasks. This led to differences when some tasks are separated differently cognitively than they are separated temporally. For example, Trent *et al.*'s model mentioned four types of analysis: Network, Host, Malware, and Forensic. My subjects discussed an automatic alert-driven analysis loop and a manual behavior-driven analysis loop as two sub-processes but both include host and network analysis. Since both network and host analysis play a part in both analysis loops they are not two distinct process components but they are distinct components in Trent *et al.* since they are distinct cognitive tasks.

This same dynamic is also seen with baselining. Trent *et al.* covers the cognitive tasks associated with baselining: Terrain Characterization identifying Network Configuration, Hosts, Services, etc. and Threat Characterization identifying False Alarms, Anomalies, etc. However, they are not separated into their activity category. Instead, they are grouped with

activities like Malware Analysis under the Analysis and Synthesis category. For cognitive analysis this grouping makes sense since the cognitive task the analysts are carrying out is similar in both baselining and malware analysis. However, in trying to describe the process used by the subject TH teams, baselining must be included as a distinct step in the process. Multiple subjects emphasized the importance of a separate block for baselining as distinct from the analysis portion of the process.

Second, tasks that Trent *et al.* included were not included in subject descriptions. 16 of the 31 codes derived from the Trent *et al.* model were either not mentioned at all in interviews or only mentioned by 1 subject. These codes included: network configuration change, host or network scan or rules of engagement. Many of these tasks map to other CPT missions like an IR or assessment missions. Additionally, 21% of the process components mentioned by my subjects did not fit into any Trent *et al.* code.

4.1.3 The Inductively Observed Threat Hunt Process

Since neither Trent *et al.* model nor TaHiTI seemed a good fit, a process model was induced from interview data. We combined the subjects' process diagrams to create a unified process in Figure 4.1. First, all nodes from all subject diagrams were placed into one interconnected diagram. Similar nodes were combined and when possible more precise nodes took precedence over general nodes. If a node was only on one diagram and was not mentioned in multiple subjects' interviews, that node was removed. In addition to the process diagram, the themes related to the process are enumerated below.

4.1.3.1 Types of Diagrams

Although observing overlap between subject diagrams and descriptions was common, completed diagrams were divergent. Four subjects provided very detailed diagrams with over 12 nodes while three subjects provided diagrams with four or fewer nodes and one subject did not feel they understood the process enough to draw a diagram at all. To discuss the difference in detail the diagrams were grouped into two categories: *Detailed diagrams*

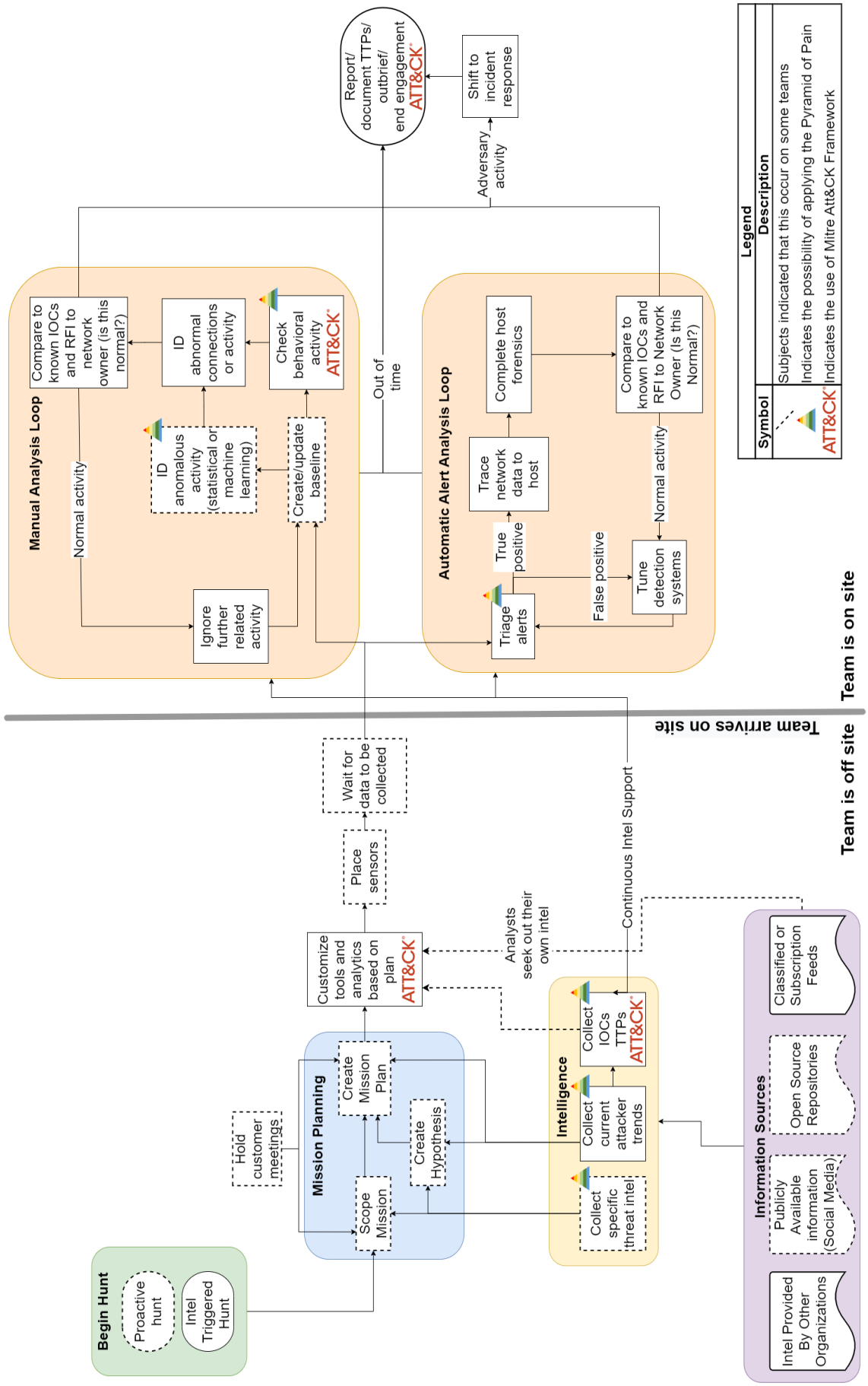


Figure 4.1. Diagram synthesized from subjects diagrams and interviews.

and *Linear diagrams*. Paragraphs on these 2 types of diagrams and on other themes that were observed are below:

4.1.3.1.1 Detailed diagrams

Five subjects included diagrams with loops and/or decision trees. We termed these diagrams, detailed diagrams. A representative diagram for the detailed diagram group can be seen in Figure 4.2. With the only exception being one subject who prepared their diagram in advance (all others drew it during the interview), diagrams with seven or more process-related nodes were created exclusively by experienced subjects (more than three years of TH experience). Unsurprisingly, the experienced members had the best understanding of the process. One subject even said: "*I guarantee if you talk to some of my more senior team members, I bet they'd have a much more clear and precise description of what we do*". Other distinctions like rank, job role or organization did not seem to have a large effect on the detail of the diagram.

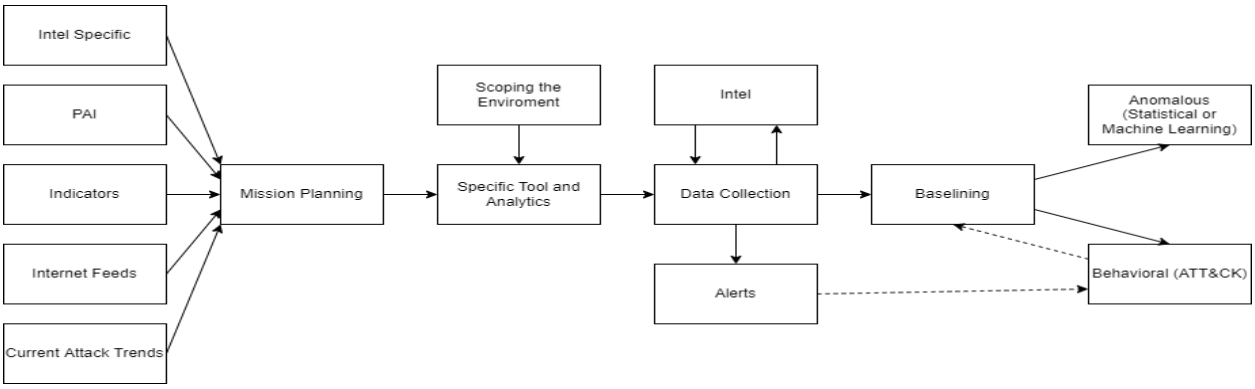


Figure 4.2. A representative diagram for the Detailed Diagram group. This group's diagram included branches and/or loops.

4.1.3.1.2 Linear diagrams

The less detailed diagrams tended to be linear and almost describe a timeline of team activity. The diagram that most resembled a timeline can be seen in Figure 4.3. The reason for the simplistic diagrams is not always related to a lack of understanding. As can be seen in

Figure 4.4, one subject drew a diagram with three nodes. This subject was a senior analyst and when describing their diagram they said "*Then, for the hunt section, the reason why it's just a nebulous cloud is because that's kind of what it is right now. We just hunt.*" It was clear from the rest of the interview that this subject knew what the typical hunt timeline looked like but felt that as an analyst there was not much structure to the tasks they were being assigned. Analysts tended to be lower-ranking personnel with higher-ranking personnel typically assigned to be the team leads. The highest-ranking personnel that were interviewed were members of leadership. Rank did not seem to have a large effect on diagrams.

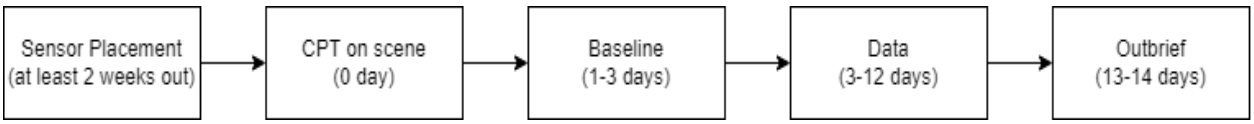


Figure 4.3. A representative diagram for the Linear Diagram group. This group's diagrams often represented a timeline of an engagement.

Figure 4.4. The least complex diagram.

4.1.3.2 Mission walk-though

This section walks through Figure 4.1.[3]

4.1.3.2.1 Begin Hunt (The Green Box in Figure 4.1)

Subjects described two ways to begin a hunt. A proactive hunt is one where the customer wants a TH team to look at the network despite the fact that there is no suspicion of adversary

[3]↑TH teams from different organizations had different names for individual hunting operations. For simplicity, I refer to them all as *missions* since that is the term most subjects used. For third-party TH teams, a mission was one TH team's engagement with a specific customer, typically lasting a few weeks or longer. For internal TH teams that did continuous hunting, a mission was delimited by a more specific hypothesis and lasted only a few days.

activity. One subject described it: "*Some of our [missions] have been ones where people just say: 'Well, we're interested in having you come see if anyone's here, but we don't have any reason to believe that somebody is there already.'*"

Triggered hunts occur when there is a reason to believe an adversary may be undetected on the organization's network. For example one subject said: "*Maybe if you're a company ... you're part of an ISAC [Information Sharing and Analysis Center] organization or another information sharing organization where you get specific intelligence that says, Hey, based on whatever data we have, we think that a system that belongs to you, may be communicating to a malicious command and control infrastructure. And that would be an example of specific intelligence where you now have something that you can say, okay, we're going in with the hypothesis that like, we are compromised specifically.*" The example given by the subject would be different than an incident response as the intelligence has not yet been confirmed. Once an intrusion is confirmed the IR team would take over or the team doing the hunt would shift to a IR mission, if they have that capability.

Triggered hunts may not always be this specific. There exists a grey area between proactive and triggered hunts. For example, an organization may know a specific APT is targeting related organizations and this APT campaign may "trigger" the request for a hunt, even though it is still being done proactively. Some teams chose not to engage in purely proactive hunts but instead want every hunt to be triggered, even if by vague intelligence such as the example just given.

4.1.3.2.2 Mission planning (The Blue Box in Figure 4.1)

Subjects reported 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. "*The next big thing, big step that we do is scoping. 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 flowing through their networks, stuff like that. So we have a scoping questionnaire we can send to the partner.*" A large organization may request a TH only on a specific part of their network. For example they

may have concerns about their financial network but not want to waste resources hunting the insecure wifi network in the employee cafe. Scoping can also be used to set expectations for what kind of threats will be hunted for. For example, time constraints may lead the TH team to constrain the hunt to only look for a specific APT vs. every possible APT and this would be agreed upon during the scoping portion of the process. Internal hunt teams may already have a defined scope of operations so may not need to repeat this portion of the process, every mission.

Hypothesis creation was described differently by different subjects and is discussed further in §4.1.3.3. One subject described it: "*we then come up with a research question. So this question is gonna be like a hypothesis almost. ... this question could be based on a threat Intel report. ... 'We believe we should not see any activity from this specific IP address because we have specific firewall rules in place, for example.' ... So that would be like a hypothesis we need to confirm or deny.*" The complexity of hypothesis changed team to team and some organizations did not use the hypothesis method at all. The example given was more specific than most external TH team's hypotheses but common for individual TH analysts or internal TH teams.

The mission plan is a formal document some teams created that planned the structure the hunt mission, almost like a timeline with associated deliverables. The mission plan is discussed in more detail in §4.1.3.2. "*mission plan campaigns are: 'what are we actually hunting for?' ... So essentially, in the mission plan, we'll include any intelligence that we have based on the scoping documentation, any previous incidents that we discovered during talking to the mission partner. ... We build that before mission. And that is used to do more like a structuring portion of the problem during the actual mission. So [the team lead] will take that, and it will be their day to day document, what they're going [to] use to track progress on the mission.*" Not every team created a mission plan document and it may be unnecessary for internal teams as their hunt missions tend to be shorter.

Subjects in leadership positions would often mention the importance of objectives in the mission planning component of the hunt. For example, one subject in a leadership position said:

> "*the really important thing for threat hunting is you go in with specific objectives. You're not just trying to find all bad activity. You wanna look for quick wins ... but you're not just going into a network to look for all bad activity forever. Right? So it's really important to define specific bounds on when you're gonna be looking for specific things on the network and what those specific things are starting on.*"

For some teams, there was a specific document that contained the objectives for a hunt. Typically, this document also included the team's hypothesis if the team implemented the hypothesis method in their hunt process. The hypothesis would serve a similar function, helping the team better define its activity.

Not all teams had formalized objectives going into a hunt. One analyst, discussing their most recent mission, said:

> "*I don't actually know what [our intel component] gave us before we went in. Oftentimes the hunts aren't necessarily specific before we start, it's more of, we know the company, we know they may or not have been compromised. ... that kind of makes it hard for us because it's like, I guess we'll just look about what's in open source intelligence and kind of go from there.*"

Other teams did document their objectives or hypotheses but did not communicate them to team members. When asked if their team used a mission plan, one experienced analyst said:

Subject: "*Yes, but I didn't have access — like, I wasn't part of it because I was just a lowly [low rank] at the time. It was the team lead who ... would come up with it.*"

Interviewer: "*Would come up with the mission plan?*"

Subject: "*Yeah*"

Interviewer: "*Did you ever see the mission plan?*"

Subject: "*Nope <laugh>It was never shared with the team so no body knew... communication is kind of an issue ...*"

Interviewer: "*What did the mission plan do if no one ever saw it?*"

Subject: "*Right. Nothing. It just made [leadership] happy ...*"

Interviewer: "*Was it counterproductive?*"

Subject: "*No, it was just never shared.*"

One team leader claimed that the mission plan is always shared with the whole team but none of the analysts that were interviewed from that organization reported having seen a mission plan.

When there is no high-level process documenting objectives or hypotheses, analysts often created hypotheses on their own. One analyst indicated they were not aware of a mission plan document, while other subjects from the organization indicated that mission plans were used and shared with all analysts. When asked about the hypothesis method this analyst said: "*Oh yeah. That's so that's what I, yeah. I try to encourage everyone to do that. I don't know if we officially write that down but, I always put that in my own notes.*" When asked about hypothesis creating and checking, another team lead said: "*we don't have a formal process for it but it's part of the normal analytic chain. So I'm sure it's done without saying.*" Three subjects total said the hypothesis method was used but not formally documented. Without formalized objectives or hypotheses, it seems unlikely that the hunt missions will be properly scoped as undocumented hypotheses might be overlooked and the hypothesis of one analyst cannot drive the scope of an entire team.

4.1.3.2.3 Intelligence (The Yellow Box in Figure 4.1)

Subjects described 3 intelligence collection tasks. (1) When a specific trigger exists, intelligence is tailored around that trigger: "*If there's already IOCs [Indicators Of Compromise] and there's already some indication of what's going on, our process is that we would take the information we have and then ... see if we can expand on, the IOC set that we've already been given.*" (2) In instances where a hunt mission is proactive, current attacker trends can be used in lieu of specific threat information: "*If there's like a current prevalent exploit that's being used, like that you're seeing in the news, like something that everyone would know about. If we have any IOCs, indicators of compromise, we're gonna go ahead and look for those. So that one's like pretty simple, we'll feed that into our tools ahead of time.*" These trends can also be used in addition to more specific intelligence. (3) Regardless if the

hunt is proactive or triggered, teams will often collect and upload additional IOCs to achieve increased coverage: "*we'll still load IOCs, right from all the threat intelligence providers.*"

Subjects agree that Cyber Threat Intelligence (CTI) is a vital aspect of TH but they differed in how their teams interacted with provided intelligence. Six subjects included "Intel" in their diagrams and two subjects noted multiple Information sources as seen in Figure 4.2. They discussed how important this intelligence was for a successful mission since it prioritizes the adversary Tactics, Techniques and Procedures (TTPs) the team needs to look for. Among the subjects, those that were analysts agreed that intelligence was crucial but had significantly more criticisms for how it was integrated into the process. See §4.2.4 for critiques.

4.1.3.2.4 Information Sources (The Purple Box in Figure 4.1)

Subjects provided many example intelligence sources but 4 sources were reoccurring. (1) A distinct intelligence team providing the TH team with intelligence: "*So we start with Intel*[4] *for threat hunting, Intel, you know, gets all the Intel <laugh>. So we have the [intelligence component] gather the Intel for us before we go on a mission ... they give us APTs and any other things that have happened to that [organization being hunted] in the past*" (2) Many team supplement that with Publicly Available Information (PAI):"*Think about the really common like social media and also just information feeds that make up the internet*" (3) Open Source Repositories:"*We'll just look about what's in open source intelligence and kind of go from there. And so that's kind of how we gear Intel to say, what's going on in the wider range of cybersecurity space during this time. And that let let's focus our efforts there*" (4) Classified or Subscription Feeds: "*We kind of gather Intel what we would call pre-mission about: what the company says happened, what we're seeing out in classified space happened, what we're seeing around us ... to be best prepared before the actual mission kicks off*"

[4] "Intel" is short for intelligence, not the tech company.

4.1.3.2.5 Customer Meetings

The mission planning task requires coordination between the TH team and the organization or customer hosting the TH team. This process typically starts by a request from the organization, sometimes followed by a survey and then meetings, typically held remotely. "*We do multiple technical phone calls with them to follow up, to discuss any details.*"

4.1.3.2.6 Customize Tools

Before the TH team begins missions, they often would fine-tune their tools using the Intelligence they collected as well as the scope, hypothesis and mission plan. "*So you plan the mission, which is just dates, what your goals are, your high level goals. And then you'll push that into developing your specific tools and analytics. So if you're hunting yourself, you probably already have most of your sensors and tools in place. You might need to attach some additional sensors though ... You might wanna stick a network sensor deeper down in that internal network, or you might need to deploy some endpoint based monitoring that gives you additional capability past what you might already have but ... 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.*" As mentioned by this subject, this task as well as sensor placement can sometimes be avoided by internal TH teams if they already have the capabilities they need to accomplish a successful hunt mission.

4.1.3.2.7 "Place Sensors" and "Wait for Data Collection" (Standalone Boxes in Figure 4.1

Sensor placement occurs when an external team would like to bring their own sensors or if an internal team needs to set up more network devices in order to have the appropriate level of coverage. "*we send a small team, maybe two, three analysts ahead of time. And their job is just to do pre deploy the kit. ... [this] gives us a baseline. So when the team shows up on site ... they already have two weeks worth of data collected.*" Not all teams performed their sensor placement in advance. This is discussed in §4.2.2.

4.1.3.2.8 Manual Analysis Loop (The Top Orange Box in Figure 4.1)

Two different modes of analysis were described by subjects (see §4.1.6). Both modes included cyclical tasks and I term them "loops". The first loop is the Manual Analysis Loop. In this loop, TH team members perform manual analysis of the collected data looking for anomalies or potentially malicious behaviors.

> "*The [manual] path I see kind of starts with understanding the environment with is some baseline analysis. ... A lot of that baseline analysis, which that path then feeds into your two like core detections, which are: you have your behavioral analysis, which is linked back to Mitre ATT&CK, and that should be driven from your primary objectives, and then you also have anomaly based analysis, which is more statistical or maybe machine learning based or algorithm based where you're looking for stuff that might not match a specific behavior, but maybe it's connected to weird things that we know a specific attacker tool uses, or it's just things that stick out from that baseline that you then can dig into deeper.*"

4.1.3.2.9 Automatic Alert Analysis Loop (The Bottom Orange Box in Figure 4.1)

The second loop is the Automatic Alert Analysis Loop. This loop is concerned with triaging notifications that the sensor automatically produces when certain IOCs are observed.

> "*It's kind of just split between two different paths on the diagram. It pretty much goes into two separate cycles ... [describes the Manual Analysis Loop] ... on the other end, it's really just triage and alert incident management at that point. ... It's really just a lot of clearing the alerts, prioritizing alerts ... If alerts are working, if we're getting true positives from there we can kind of drill down and do our network and host forensics, find the root cause, find if it's network traffic: 'Where's the network traffic coming from? What host is it originating from?' If it's a host alert, you know, pulling an image, and*

digging down to find what exactly did we alert on? And once we get that start to finish of what happened: 'Why did it happen?' Then we go to the network owner and start to figure out: 'Is this something that we've just been chasing our tails on? Or is this something that should not be happening?' From there, that's where that the cycle on that side comes from is just triage. The alert, [then] investigate 'is this normal?' And then reassess, and just keep going back through those [alerts]."

4.1.3.2.10 Shift to Incident Response

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. Some TH teams perform multiple functions and are also capable of acting as IR teams: "*Let's say we have a discovery, we discovered something. The next step would be, ... depending on what we find, we may need to turn the hunt into an IR. So it becomes more focused on only on this specific threat that we are going to concentrate on, and help [the customer] to remediate and mitigate depending on security.*" Other TH teams may not be able to "*turn the hunt into an IR*" and instead they would need to call a separate IR team. It is also worth noting that not every finding needs to trigger a hunt. This subject continues: "*If we find commodity malware on the computer, it may not require a lot of resources. Maybe we'll dedicate one or two analysts just concentrating on this specific issue. The rest would continue with the mission plan. If it's something bigger ... we're probably going to shift into IR mode.*"

4.1.3.2.11 Report/Document TTPs/Outbrief/End Engagement

At the end of the mission, TH teams typically write a report. This report documents all findings. It will often provide recommendations. It is sometimes presented to the network owners.

"*We come back to those hypothesis in our final report, too. So we come to the final report and then each one of those hypothesis, we do a certain amount of investigation into that hypothesis by doing those curated analyst queries.*

And that is something that [the team leads] document in the report at the end, which 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. So, sometimes obviously we find things that are not related to our hypotheses because we still do queries that aren't only related to those hypotheses. Right. So if we go in a different direction, than our final reports document those additional findings as well."

4.1.3.3 Analysis Frameworks

The hypothesis method is a common method used in Threat Hunting [37][90][9][78]. When asking about hypothesis documentation a few two subjects from one organization (one in leadership and one team lead) mentioned a written document called the mission plan. For some teams, this plan is created before the mission and contains 1-3 high-level hypotheses along the lines of: *APT #100 exploits VPN device x.x.x.x and pivots to the Domain Controller.* Out of the three organizations that were studied, two formally documented their hypothesis (one in the mission plan and one elsewhere) and one organization did not include hypothesis documentation as a part of their process.

In addition to the hypothesis method, each subject was specifically asked about three other models observed in TH descriptions across academia and the private sector (see §2.2.1). The results can be seen in Table 4.1. Across all three organizations, Mitre's ATT&CK framework was the most used although the role it played in the process was not the same for every subject. One subject discussed how it was used only as an aid in describing behaviors written up in the final report. Other subjects claimed it was used to categorize tasks assigned to analysts and others claim only the intelligence component of the process used ATT&CK to categorize indicators and detail adversary behaviors. Kill chain was only used as far as it maps closely to the same categories in the Mitre ATT&CK framework: "*Our processes are loosely based on both [ATT&CK and the Kill Chain]. The actual detailed hunting process is very closely related to the ATT&CK framework.*" The Pyramid of Pain model was also not

Table 4.1. Frameworks incorporated by subjects' process.

Analysis Framework	**# of subjects who claim their process incorporated the framework, by name or informally (# orgs)**
Mitre ATT&CK	11 (3)
Hypothesis Checking	8 (3)
Lockheed's Kill Chain	3 (2)
Pyramid of Pain	2 (2)

frequently used, although some subjects claimed to use it implicitly in their work and one subject indicated it was their goal to achieve more widespread usage of that model.

4.1.3.4 Standardization

Across all three organizations, subjects claimed everyone within their organization used the same process. Three subjects even claimed the process would look similar outside of their specific agency but within their government department (DHS or DOE). This claim was also paired with statements indicating the process was very flexible. Other statements indicated that the process was even more flexible. One subject discussed how a spreadsheet was used to track the process (a popular tracking mechanism according to subjects) but that the team lead could decide what gets placed on the spreadsheet. Another subject said that the team lead could choose not to use the spreadsheet at all. One team lead shared the sentiment in this quote: "*there is a checklist that kind of went over the steps I alluded to but I would [say, in] my own personal opinion, the steps are useless and nobody actually followed them. That was good, so the teams were going out and finding stuff did all those steps in [their] own analysis, regardless of them being part of the checklist but the checklist was there for someone to ensure minimum quality standards.*" This seems to contradict this subject's own statement earlier in the interview that their organization's process is standardized with their organization. Contradictions like this indicate that processes being used within an organization may vary team to team.

4.1.4 Process Creation

Process creation differed from organization to organization. One organization's process was created by a group of team leads. At this organization, the team leads seem to still be the primary maintainers and reviewers of the process. At another organization, the process was created by 1-2 people in leadership positions who purposefully imitated other government teams. This imitation allowed the organization's teams to interoperate with other government organizations.[5] After its creation, the process has been reviewed and edited by a spectrum of maintainers. One subject claimed that "*everyone who had prior experiences or is now part of the [team] and been on a mission, have access to our [process document] to submit edits*" and another subject (an analyst) claimed that their suggestions were incorporated into the process.

4.1.5 Process Changes

The TH process at all three organizations changes frequently. Eleven out of twelve subjects said their process was in a constant state of change. However, no subject indicated that this was problematic. Instead, many subjects claimed the fast turnaround on updates allows issues to be dealt with quickly. For example, when asked about how static the process is, subjects responded: "*It's changing constantly...every time you do something you should learn and update how you're doing it.*" and "*That's the nice thing about building the whole process out. Because we are so new ... 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 subject warned that process change on site 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.*" However this team lead also said they implemented analyst feedback, changing their process from one mission to the next: "*They would say, 'Hey, we wanna try this out this hunt'... or, Hey, we wanna do this. So I would just kind of listen to what they want and occasionally, I would*

[5]↑Interoperability is desirable because a large scale cyberattack may require organizations to lend cybersecurity personnel to other agencies.

interject my own things... like, Hey, I really need you to do this, this time but more often than not, I would let the analyst decide." Updating the process, even frequently, (as often as every mission) seems to be helpful as long as it is not changing while on mission.

4.1.6 Automated Alert Loop vs. Manual Loop

Figure 4.1 includes two analysis loops: an automated alert loop and a manual analysis loop. 2 subjects, one team lead and one in leadership, specifically discussed the two loops as independent. Adversary activity was found almost exclusively during the manual loop. Out of the subjects that had detected adversaries on hunt missions (7/12), in none of their examples was the activity found by the alert loop. Four subjects had only discovered adversary activity on one mission but all four said that the activity was discovered by an analyst, not an alert. Out of the three subjects that had multiple hunt missions where activity was detected, two out of three said that activity was primarily detected by analysts. For example, one team lead said:

> "*I worked mostly against advanced present threat actors. ... it was not very often that it came from an intelligence report and it did not often come from automated sensing. It was almost exclusively found by analysts observing the data for living off the land techniques or new zero days.*[6] *I would say that we found at least three or four zero days, which, couldn't have been detected by the other two methods. Well, you could technically do it via anomaly detection, but, in practice, it never really happened.*"

The second subject, a member of leadership, said: "*A lot of things we find are mostly analytical kind of behavioral activity. Like an analyst spots. I'll just speak the most recent...*" and the subject goes on it describe how an analyst found activity. The third subject did not comment on this question.

[6]↑"Living off the land techniques" refer to adversaries making use of tools already installed on the computer instead of software which would have to be downloaded to the device. "Zero day" refers to a vulnerability that has not yet been discovered by network defenders.

Table 4.2. Process issues noted by subjects, with their proposed solutions and related open problems. This table, and the prose, only includes issues noted by ≥ 4 subjects across ≥ 2 organizations.

Observed process issue	# subjects (# orgs)	Proposed Solutions	Open Problems
Too little automation (§4.2.1)	7 (3)	Automate baselining and equipment set-up	Will automation hinder the analysts? Automation management?
Teams lack needed data (§4.2.2)	5 (2)	Deploy sensors ahead of time	What data to collect?
Insufficient process detail (§4.2.3)	5 (2)	Add SIEM queries	Make the process more specific while maintaining process flexibility?
Poor threat intel (§4.2.4)	5 (2)	Quality indicators with broad coverage	
Inaccurate task tracking (§4.2.6)	4 (1)	Make team leads track task completion	
Side-tracked analysts (§4.2.5)	4 (2)	Define clear objectives	Identify rabbit holes early?
High turnover (§4.2.7)	4 (2)	(1) Pair newer and experienced members; (2) Enhance process documentation	

4.2 RQ1.2: What shortcomings exist with current government Threat Hunt processes and what can be done to alleviate these shortcomings?

This section describes six TH process shortcomings beginning with the issues mentioned by the largest number of subjects and proceeding in descending order. These are summarized in Table 4.2. Only issues that had four or more mentions were included.

4.2.1 Automation

4.2.1.1 Problems

All three 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%*". One subject claimed 50% of their process was automated but the automation was geared towards data visualization. They explained: "*we don't have any true automation*" indicating that the data analysis was done mostly manually aside from alerts and a simple script that assists in baselining.

Some teams had automated queries in dashboards, kit setup scripts, and alerting. Other subjects' claimed the only thing automated for their team was what the vendor set up in their tools. Multiple areas that could be automated were brought up by subjects and discussed below.

Seven subjects indicated dis-satisfaction with the current level of automation. They made statements like "*not as much as we'd like*" and "*that's a big point that could be improved*". Three additional subjects indicated that the process was all or mostly manual, making statements like: "*It's still pretty manual.*". One subject described what was automated and provided examples of things that could be automated but did not seem dissatisfied with the current level of automation. The last subject was not asked about automation due to the interview's time constraints.

Eight subjects described hindrances to getting more of the process automated. Three subjects mentioned that analysts were not given a sufficient amount of time to generate automation. When asked about hindering factors these subjects made comments like "*It's always time, right?*" or "*Mainly just lack of ... development time. We've just been very busy.*" Three subjects described a lack of personnel. For example, one subject described a situation where a contractor had been working on automating components of the process but "*he just left.*" Another 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 no one understands [tool]*". The other subject said: "*Just ex-*

perience. Knowledge and experience. We're fairly new to all of this so having that baseline knowledge ... from a mature organization is isn't there. We will grow into it, but having people thrown in from all over the place with very little knowledge kind of hinders the whole process [of automation]." Other hindrances that were mentioned only by one subject (not always the same subject) included not having enough money to purchase good automation solutions, automation solutions not being fast enough for short two-week engagements, and the organization having other priorities that conflicted with automating tasks. One team lead also described that their team only reported to them while they were on a mission. At other times they did not have control over the analysts and could not task analysts with automating tasks.

Although most subjects believed more could be automated, there was less agreement on how much of the process could be automated and what tasks should be automated. How much more could be automated range across subjects from "*40% of hunt tasks*" to "*70 to 80 [%]*".

Two subjects hesitated about further automation. One subject, a team lead, had a psychological concern. They thought analysts should be making all the automation they use, themselves. They said: "*I would say you should not have automated things available to the teams as part of the [hunt process]. Each team would need to build their own automations for their own style and technique. If you provide people automated analysis ... it will become a crutch that they will not do analysis on their own because: Hey, I ran all the automated things. I didn't find anything. Let's close the report.*" They also indicated that no analysis should be automated.

The other subject, a member of leadership, had a philosophical concern:

> "*I think we have to maintain a balance, because I think hunting in general should start where automation stops. That's like the whole premise of hunting, right? Because you have these SOC automation tools, that you're supposed to monitor, and do all these things, triaging, and identifying, and bubbling things to the top for you but with threat hunting, the premise, the fundamental assumption, is that the sophisticated adversary already bypassed all that and*

now you have to apply the manual technique to be able to find them. So I don't know, I think to some point it's good to automate some things that are repetitive, maybe like deploying kit. But I think most of the analytical work should still be relying on human factor."

The idea that Threat Hunting cannot be automated is a common concern in TH literature [34], [69].

4.2.1.2 Solutions

Five subjects suggested areas that could be automated. Two subjects said baselining took too long and that it could be almost fully automated, especially for less complex environments. One of them said: "*I think the baselining could be fully automated, to a point. At least more automated than it is right now.*" Two subjects discussed equipment preparation tasks. One suggested configuring and updating the software on the TH equipment could be automated: "*as far as the build goes, it could probably be [automated] quite a bit. Probably 80% of it.*" The other suggested automating the process of uploading alerting rules for indicators of compromise (IOCs): "*It would be cool to automate the indicator portion, other than us having to manually feed in that stuff.*" These "indicators" needs to be refreshed and re-uploaded every mission. Individual subjects also suggested (1) automating the communication of important artifacts and their contexts to the team, (2) automating the gathering of logs from endpoints and (3) making analysis repeatable by having analysts write automated scripts that perform their analysis as they go.

Although not an automation suggestion, one subject discussed how their team had plenty of automation for their tasks. Their primary concern was creating a centralized location for these automations. They said: "*We're kind of crippled by how little we have in terms of preconfigured tools, scripts, whatever it is.*" They later explained: "*We're not great at managing our central repository of automated tools... Mostly because we've changed our method of repositories like three times in the past year so that's definitely something that's lacking in automation. There's plenty of people with ideas, plenty of people with finished*

products, almost finished products on how to do those [automation], but we're not great at disseminating it [or] storing it."

4.2.2 Sensor Placement and Data collection

4.2.2.1 Problems

Many subjects indicated the process was counterproductive in the way it collected data. 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 data had been collected. Without data, threat hunting is virtually impossible. As one experienced subject said: "*you can't find anything without data*".

Seven subjects either mention this issue (2) or said that they deploy sensors ahead of time (5). All subjects that mentioned deploying sensors early gave this issue as a rationale. When asked for an example of a hunt mission where the process was counterproductive, two subjects indicated that the sensor placement at the same time as the team arrival was a hindrance to the mission. For example, one subject said "*We did a threat hunting engagement ... we were focused on the problem for about two weeks, but we plugged sensors in on day one of those two weeks. So when you do that, you don't have a baseline of what even one work week looks like much less, a couple work weeks. And, you know, the cycles that every network goes through over the course of days, weeks, months. So that was not very effective [because] we didn't have the right data we needed in the baseline to actually do real analysis on the network.*"

Another data collection issue is that some teams collect too much data or the wrong type of data. This extraneous data obscures potentially relevant data. One experienced subject said: "*I know with [former TH organization], ... back in the day we would ask for a ton of stuff. Give us this, and that, and that, and that, and for no reason, right? The way we do it here, we rely on our own tools in the beginning, and intelligence. ... 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 provided two examples of previous missions where there was a mismatch between the data provided to the team and the expectations of the process. In one example too much data was provided and 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 ... having the checklist and trying to stick to it was counterproductive because if you tried to stick to it, the data made less sense than if you didn't have a checklist.*" In the other example this subject provided, the team had the opposite problem. Due to "*not a lot of logging*" and inadequate log retention, the team was not able to trace adversary activity from initial exploitation to the present day; the process assumed this would be possible. The subject reported: "*we found...indicat[ors] of threat actor activity, but...no common theme. We could only prove discrete things ...we couldn't find [things] that the checklist was assuming we could.*" The amount and type of data that is collected, and when it is collected affect the success of a hunt mission. Many TH teams find it challenging.

4.2.2.2 Solutions

Five subjects had been on teams that had experienced the issue of not having data when the team began analysis. These subjects described a solution that worked for their teams: A predeployment team would install sensors at the location being hunted, multiple weeks before the teams' arrival. This ensures data collection begins before the arrival of the TH team. One subject said: "*the big change that I've seen made in my time at the [hunt team] is the sensor placement in advance. That is something that we implemented since I reported. That's been helpful to our [hunt team].*"

Interviewer: "*Awesome. And what caused that change?*"

Subject: "*What caused that change was having [hunt teams] go on mission and having to wait a few days to get data before they could actually start. We wanted them as soon as they got on scene to be able to jump right in so it was not [a] waste [of] time or money.*"

Another subject agreed:

> "*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? Like the setting sensors in advance. Our first time that we did this, we just set sensors the day that we arrived and that was terrible. <laugh>basically, it took the whole first week to create a baseline and we didn't even have, a high level of confidence in that baseline because that data was just actively being gathered. Now with at least like the two weeks of data that we've gathered, we can pretty much see patterns and stuff just like from week to week, at least. So that was a huge improvement. It took the baseline portion of our mission plan down from like the whole first week, basically to just like the first two and a half days*"

The issue of unhelpful data being collected seemed to be common. No subjects mentioned having solved this issue. One experienced subject suggested more precise scoping as a possible solution. They said: "*we go into these environments, you see a lot of guest networks, or IOT [Internet of Things] devices, like cameras and security badges and stuff. It's stuff that we don't necessarily need to monitor, unless we have a reason to, but by default we're just like, grab it all. Why not? And then it just fills up our sensors. And then you run a query, and then it takes forever to get results. And there's much data, it creates lots of noise in this. Instead of just concentrating on like, what is your server subnet? Or what is your DMZ? Let's just figure out, let's concentrate on one thing, but do more in depth work, concentrate on quality, I guess, versus quantity.*" Other subjects spoke about the importance of scoping, but never specifically connect it to the issue of too much data. Scoping was discussed in §4.1.3.2.

4.2.3 Process Documentation

4.2.3.1 Problems

The level of detail in process documentation varied by team. One organization had a detailed checklist but it was not used. Another organization had a similarly granular process but allowed team leads to make changes if necessary. A member of the leadership from this

organization said "*[team leads] are responsible for their mission. So they are given leeway to adjust as things happen.*" When asked how often this occurs this subject said: "*I'd say 25 - 50% of the time, but usually that's just because they found something they want to investigate and they may prolong the investigation, which pushes investigating something else to another day*". The third organization only had a high-level process. One junior analyst from this organization was concerned about the lack of repeatability: "*we all have a very, very, very different workflow of how we actually go about analyzing things but I think in the future if we can work on formalizing it... I think that will help for repeatability.*"

A desire for more process documentation was common across five subjects who asked for better process definition, more specifics, or more coverage to be added to the process. Only one subject said the 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 that was useless because the customer did not use VPN] so like there's some things that just like it's too specific, you know? And I think it's almost better, you know, if we would like make that more abstract and say like: 'Hey, why don't [you] just look for any remote access software?'*". This analyst later requested additional process information be included for less experienced analysts.

4.2.3.2 Solutions

While one subject explicitly said the process was too detailed, a second subject gave an example of a mission where the specificity of the process was counter-productive. A more detailed description of the mission was provided in §4.2.2 but the subject was a team leader on a mission where, due to equipment issues, many of the specific components of the process did not apply. When the team lead decided to use a more abstract process to lead the team, the mission was more successful. This example is important because, as mentioned previously, a different subject requested better-documented minimum expectations. If these minimum expectations are too rigid and too detailed, a team lead in a similar situation might similarly find them to be a hindrance. One of the organizations planned for such contingencies by enabling team leads to make process changes. A subject from this organization's leadership

was asked about the rigidity of the process. They said: "*it's fairly rigid. Um, we usually try to give our [personnel] a hard cut off. The only exception would be is if it is a, um, national priority for us to go do the mission.*"

Interviewer: "*And do you think that that is a good thing, how rigid it is, or do you wish it were less rigid?*"

Subject: "*I think it's a good thing.*"

Interviewer: "*Okay. And why is that?*"

Subject: "*It allows our [nontechnical personnel] who have less cyber background to have guidance to work within.*"

Later in the interview, the subject indicated that team leads could change the process, if required by the mission. When asked if this is what they meant to say the subject responded: "*Correct. They can change it if they find something that they think warrant's changing, provided it's briefed up.*"

In this exchange, a tension between flexibility and definition can be seen. Process definition helps provide structure, which is especially important to less experienced personnel. However, in situations like the example mission where detailed parts of the process didn't make sense in the context, flexibility must be afforded to allow the team to adjust. The subject's organization attempted to balance this by allowing such changes at a team lead level. Another subject also in leadership spoke on the importance of this level of flexibility: "*I dislike making rigid guidelines. ... At the end of the day, I think the goals of each engagement may be different based on the partner, the threat actor, the geopolitics around whatever the thing is, ... if it's totally proactive, ... So 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.*" Since more detailed documentation was requested by subjects, this additional process documentation must not infringe on the flexibility afforded to the team. If it does, the process could become unintentionally counter-productive. This is especially true if additional minimum expectations are added to the process.

The most common feature requested in the documentation was SIEM queries, mentioned by three of five subjects. Two analysts and a team lead noted: "*It would be good to have*

queries written at least ...for each [task] as a starting point as to say: 'You should start with this query for this' and then from there investigate multiple different like data sources or log sources to go further down the trail." or "*it would've been nice if we had a digital bookshelf of just a bunch of preconfigured [tool] rules that we could just grab and bring with us. Show up on site, boom! Here's our prioritized queries that we have in [tool], something like that.*" Such comments indicate SIEM queries may be a good place to start.

4.2.4 Cyber Threat Intelligence

4.2.4.1 Problems

As discussed in §4.1.3.2, many subjects were unsatisfied with the threat intelligence provided by their organization's intelligence component. Five subjects indicated that the intelligence component of their organization had room for improvement. Their criticisms ranged from saying intelligence provided "*led us on too many false flags*" or "*I'll be honest with you, we're not great with intelligence yet*" to "*We have the [intelligence component] gather the Intel for us before we go on a mission. Usually, I honestly don't know how they do it because it doesn't make sense. ... It seems like they just randomly pick out APTs to go after.*" According to Bianco's hunt maturity model [69], good threat intelligence is a basic threat hunting requirement that separates a level zero hunt team from a level one hunt team. Levels range 0-4 with four being most mature (see §2.4). Teams at level zero are not considered able to perform Threat Hunting missions.

4.2.4.2 Solutions

One aspect of threat intelligence that was often brought up by subjects was the indicators provided by the intelligence component. Indicator sets can have two important features: quality and breadth. Low-quality indicators could be indicators that are old and reflect activity that is no longer characteristic of adversary activity. For example, low quality indicators may alert on IP addresses that are no longer used by adversaries. A narrow set of indicators could be indicators that concentrate on a type of compromise or specific actor.

For example, indicators that are focused on email server compromises, may not alert on a VPN server compromise.

The only subject who explicitly discussed the issue of quality vs. breadth was an experienced member of leadership who favored breadth of indicators. They said: "*quantity over quality right? Quality is important too, but the more indicators you can cover and the broader the surface, the more likely that you'll get that quick win from finding the one [Command and Control] server that happens to be like a workstation or servers beaconing back*". This subject went on to describe the importance of filtering alerts properly: "*you need filtering on there so your analysts aren't getting overloaded, but you wanna catch as much as possible and then kind of funnel it*". The subject goes on to say how important breadth is for an indicator set. In this quote, the subject also mentions that analysts can get overloaded with too many indicators. Overloaded analysts will miss things and this could be another reason why alerts are less likely to pick up adversary activity. The subject complaining about their Intelligence component picking random APTs (quoted in §4.2.4) likely had a similar concern that the indicators they were given did not have enough breadth.

Both breadth and quality of indicators are important. One team lead indicated that the intelligence they were provided led to too many false positives. False positives are typically characteristic of low quality indicators so it seems this subject valued quality over quantity. A member of leadership bragging about the indicators they were using, said: "*on the most recent hunt mission that we did at [location], we implemented or ingested in [tool], close to 700,000 indicators, recent indicators too, not stale.*" This subject valued both breadth (700,000) and quality (not stale). Quality and breadth do not necessarily conflict but a broad collection of quality indicators may be difficult or expensive to acquire.

4.2.5 Side-Tracked Analysts

4.2.5.1 Problems

The issue of analysts spending too much time on one task was mentioned by four subjects, two team leads and two analysts. Typically when this occurs the analyst goes into such detail trying to prove that something is not malicious that they lose sight of the bigger picture.

Such a line of investigation is often sufficiently unlikely that other, more probable scenarios, should be investigated first. When an analyst does not check in with their supervisor for a long amount of time because they are engaged in investigating something that is not likely to be productive, this investigation is called a "rabbit holes" Both team leads mentioned the rabbit hole issue. The two analysts noted it could be problematic depending "*on the person*" or without proper objectives. The second analyst said: "*[The hypothesis] actually helps focus me and stop ... the rabbit hole effect... I think having the hypothesis helps keep you focused and on task and helps you sift through the results that are unimportant.*" No other strategy for rabbit holes was proposed.

One analyst was annoyed that leadership stopped them from investigating rabbit holes. They said "*I really like to dive deep on something if I find something interesting. I like to spend a lot of time like researching it and because ... the [process checklist] is so long sometimes they will rush me and say, 'Hey, don't look at that.' You need to go ahead and finish up the [checklist] and stuff like that.*"

4.2.5.2 Solutions

It was unclear from the interview data how to balance in-depth hunting against the risk of rabbit holes. One subjected mentioned above indicated that the hypothesis assisted them in not going down rabbit holes but no other mitigation was discussed by the subjects. See §5.2.7 for further discussion of this open question.

4.2.6 Process Tracking

4.2.6.1 Problems

When a team is on-site, the team typically uses a system to track the progress made on different process tasks. This system led to an issue of analysts "closing out" tasks before they had been fully investigated. This issue was mentioned by four subjects. Two subjects provided examples from recent hunts where this had been an issue. One analyst said "*I can think of several times in the last mission where there was a task that was checked off and*

then I went back through it and looked through it and I was like, this is definitely not fully investigated."

4.2.6.2 Solutions

To avoid incomplete tasks being marked as accomplished, some team leads did not allow analysts to close out tasks. Theses team leads made process tracking a task for the team leader. One team lead said: "*I keep it at myself or [another lead] to check off tasks cause a lot of times you'll have an analyst that does a task 20% of the way through and marks it as complete.*" A different experienced team lead made a similar decision to track items, however their motivation was to make things easier for the analysts. They said: "*I felt the best system was to take the burden off the analyst so they could just kind of chase what they were doing. [Describes the communication between an analysts and a team lead] and then it was up to the team lead to translate that into what it meant on the checklist. Anything that was pushing it down to the analyst to make the team leads' job more convenient, I think hindered the analysts because if you were forcing them to open tickets some of them would either not open the tickets or if you asked them to, or if you force them to, they would just not do as good of work, because then they were just like, I don't wanna do this. This is tedious.*"

4.2.7 Turnover

4.2.7.1 Problems

§4.4 discusses process changes that could assist with integrating new members into a TH team. Four subjects mentioned turnover negatively affecting the team. For example, 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*". Another subject said their entire team was getting replaced at once with new inexperienced members.

4.2.7.2 Solutions

Multiple possible solutions were offered by subjects - pairing new members with more experienced ones (§4.4.1), assigning novice members tasks that are compatible with their abilities (§4.4.2) and improving documentation for new members (§4.4.3). No solution was decisive but the literature [18] does indicate that good processes can minimize the impact of inexperienced teams or teams with high turnover. Recommendations discussed further in §4.4.

4.3 RQ2.1: How do newer members fit into the Threat Hunt process?

The sections recounts answers to two lines of questions: (1) questions regarding what tasks newer members were currently performing in TH teams and (2) questions inquiring what variables that were affecting the time it took these newer members to become integrated team members.

4.3.1 Tasks Assigned to Newer Members

Three subjects spoke about dividing up threat-hunting tasks based on experience. One subject said: "*we have a huge list of a variety of indicators and some of them include... and we kind of take that within our team and assign it out to people mainly based on experience*". They described the simpler tasks being given to newer members: "*The [tasks] that are easier, like top 10 DNS requests for the network, we'll give that to some people who are just kind of like learning*". Another subject described appreciating having a large number of newer members earlier in the engagement: "*In the early days, it is great to have just a lot of eyes so having junior members is a good thing because when you're looking for the low-hanging fruit. It's just about more eyes on target.*" This subject goes on to say that experienced analysts are more useful but: "*even though the junior analysts were not the ones finding the things, they were very useful at proving the negatives. So once you had a thread to pull, you could task them with: 'I need you to tell me all these things that would take a lot of time, but they're very discreet tasks' and that would free up the senior analysts to focus on the*

more fruitful things. ... so they were good at proving the negatives, which allowed the senior analyst to focus on finding more things. So I put a lot of the busy work onto the junior analyst which is normal."

In addition to being "*good at proving the negatives*", three subjects mentioned that new members provided a helpful outside perspective. For example, one subject said: "*It's kind of beneficial to talk to someone who's not within the community to ... see whether a process makes sense, see if the training makes sense, ... [to] question you, ... 'why are you doing these steps?'*". They later described a newer member creating a useful technique that helped the team look for certain adversary behavior.

One of the team leads believed that newer members could sometimes be seen as a hindrance. They explained: "*The junior analysts, uh, often impede everyone because they are just asking more questions and they are taking time away from these senior analysts.*" This team lead described an example of a time where this was observed: "*So in this case I kept [the novice members] on for a long time but at some point when the entire team was very frustrated, I fired the rest of them. That way the senior analysts could focus.*" At the end of the interview they added that since their experience was on a TH team targeting APTs, it was not a good place to integrate junior analysts: "*Everything I've spoken to, it's mostly advanced persistent threat. This is a horrible place to integrate junior members ... You should not integrate people, junior analysts into a team like this. And if you are, you need to be able to make them understand what's going on.*" This lead's views were unique in two ways, they provided more examples than most other subjects regarding challenges with newer analysts and they also seemed to view newer analysts more negatively than any other subject.

4.3.2 Factors That Affect Time to integration

To understand how long it takes a new member to become integrated into the team, the interview instrument included: (1) "*how long does it take for a new member to operate independently?*" and (2) "*how long does it take for a new member to become a net positive for the team?*". There was no consensus on either answer. Timelines ranged from three months to one year to become independent and from two weeks to two years to become

Table 4.3. These factors reportedly effect new member integration time.

Factors that effect new member integration time	**# subjects (# orgs)**
Computer & Networking Basics (§4.3.2.1)	5 (1)
Cybersecurity Education & Experience (§4.3.2.2)	5 (3)
Number of missions (§4.3.2.3)	4 (1)

a net positive contributor. During these lines of questions, 10 out of 11 subjects signaled uncertainty, saying something like "*I think that's gonna be a really subjective answer*", "*I don't know...*", or "*that one's a bit hard to say...*".[7] The team lead that seemed confident was the only subject in the study that had never been on a hunt mission.[8] The most interesting result of these questions were the traits that the subjects mentioned as affecting these times. These were cybersecurity education and experience of the new members, the new members basic computer and networking non-security knowledge and the number of mission the new members had been on in the new organization. These are summarized in Table 4.3 and elaborated below.

4.3.2.1 Computer & Networking Basics

Five subjects mentioned the basics of computing and networking. They made comments like "*a basic understanding of how computer systems and networks work is really valuable*" or "*honestly if I had to make one requirement for a knowledge base for anybody that comes on mission with us, it would essentially [having] the equivalent of a network architect's level of understanding of a network*". One of the five subjects said that although the basics were helpful, an analyst's mindset was even more important. Their quote is in §4.3.2.4.

[7]↑These questions were skipped in one interview due to time constraints.
[8]↑They said: "*Yeah, so usually after their first mission, they're able to take away and to know what questions to ask and then as when they're not on mission. Use that time to kind of dig deeper into things that they saw... usually by the second mission they're able to [be a net positive].*"

4.3.2.2 Cybersecurity Education & Experience

Five subjects mentioned some kind of general cybersecurity background as being helpful. They made comments like "*I'd say cyber security fundamentals and actually understanding how those systems behave is really useful knowledge*" or "*we look for ... college education within computer science, cybersecurity information assurance, information security*".

Two subjects indicated that some cybersecurity backgrounds were better than others. One subject said "*it varies based on even just prior experience in cyber security. I would say my most confident members have done incident response. I think that's a huge one because they already kind of know what bad looks like*". Along these lines, another subject pointed teammates with certain cybersecurity backgrounds would be lacking in different areas. They said: "*I think if it was a member who worked at a [Security Operation Center (SOC)], they would be comfortable with working within a [Security Information and Event Management system (SIEM)] so that would be good and also working with an [Endpoint Detection and Response System (EDR)] so using the tools, they would be fine. So that would help a lot but part of our job is also to set everything up and also if they've only been at a SOC, if they've never done any like pen-testing themselves, any offensive actions, it's also harder to hunt for stuff like that, that you've only ever read about and you've never actually done yourself.*"

If subjects did not identify with backgrounds that helped decrease integration time for newer members, they were asked about hunt experience. Five subjects mentioned either on their own or as a result of this question, that prior threat hunting experience would decrease this time for newer members. When asked about members that had cybersecurity experience but were new to the team, one of the five said: "*Yeah, absolutely. I would say someone who's coming in who doesn't have experience with [team-specific processes] can, within one mission, probably just go ahead and get up and running like one mission with some help in the next mission. They'll probably be okay.*". Another subject responded: "*Yeah, definitely... I mean the first [mission] we were on, we had someone who had been on a hunt team before and they were really helping shape that mission plan since we didn't have one at the time. They were kind of building that from scratch for us.*"

4.3.2.3 Number of Missions

Four subjects believed the time it took for new members to become a net positive or operate independently was tied to the number of missions they had gone on with the organization. For example, when asked what could shorten this time, one subject answered: "*More missions? ... Yeah. <laugh>, you know the best thing is actually doing it. Right. Okay. So that's uh, I mean, yeah, that's a huge part of it is just getting more missions outta your belt, which I guess that's not really like a good answer,*". All four subjects that mentioned the number of missions being important, based their time estimates on how many missions a new member would be going on. When asked one of the two questions about time to independence or time to be a net positive, they made statements like "*I would say really, maybe two missions*" or "*Weeks to, yeah, a little more. Well, it all depends on how many missions you have.*"

4.3.2.4 Other Factors

Many factors received less than 4 mentions by subjects. These are enumerating here:

- Two subjects mentioned the leadership of the team can affect the speed at which new members can integrate themselves. One subject said that given the correct traits an analyst could be integrated quickly, "*especially with the right leadership too because a lot of it isn't driven by ... asking the right question to lead them to the next step, not necessarily them getting there themselves.*"

- Two subjects mentioned if new analysts were comfortable with the tools they would be able to reach these benchmarks more quickly.

- Two subjects mentioned training. Their quotes are discussed in §4.5.2.

- One subject in leadership indicated that members who start their career in cybersecurity are better prepared to do the job than those that began in an adjacent field: "*We don't have a lot of senior people right now that have come up through the ranks knowing this stuff. So those in my position struggle a little bit more than the kids that*

are coming at this from a junior-level position. Does that make sense? I don't have the background pretty much as these junior kids and as they come up, they'll be able to fill my role better in a better situation, better ... light I guess."

- Two subjects mentioned mindset but had different meanings for them. One subject thought mindset was inherent and said: "*I value the mentality a lot more than I value the specific technical training... I think someone with a very basic understanding of those with the right mindset could very quickly be a good analyst.*" while the second subject believed that came from TH experience.

- One subject mentioned willingness to learn, saying "*one of our best analysts really had no background in cyber security information security but was so willing to put the time in to learn that now they're thriving.*"

- Lastly, one subject said that analysts with "*passion*" would integrate more quickly than those without it.

4.4 RQ2.2: How could process changes facilitate the integration of less expert members?

During the interviews, three themes emerged when asking about how to better integrate less expert members. These three themes as well as two minor themes are presented in the Table 4.4 below.

4.4.1 Pairing members

Seven subjects spoke about their team pairing new members with expert members as a way to improve the integration of newer members. None of the seven subjects spoke unfavorably about this practice as a method of improving integration. Two subjects specifically said it did help. One said: "*I've observed the implementation of that, and I've seen that it helps stand up ... our other team faster*". The other subject said: "*One of the things we did ... that I thought was really useful is doing a threat hunt with fellow IR member. They kind of had experience looking a lot of the data and the host forensic side things and so I was able*

Table 4.4. Subjects reported these recommendations for integrating new members effectively. *These recommendations were brought up by subjects but are not specifically process related.

Recommendation	**# subjects (from # orgs)**	**Explanation**
Pairing Members (§4.4.1)	7 (3)	Pairing new members with veteran members as a training methodology
Task Separation (§4.4.2)	4 (1)	Giving new members appropriate tasks
Process Documentation (§4.4.3)	6 (3)	Having process documentation new members can reference
Training* (§4.4.4.1)	3 (2)	Providing new members with tool-oriented or on-the-job training
Personnel Issues* (§4.4.4.2)	2 (1)	Extraneous organizational difficulties

to learn a lot from them ... so I thought that was really useful. I think like a lot of shoulder surfing kind of experiences, I think that would be useful.". Shoulder Surfing is the practice of a novice, looking over the shoulder of a more experienced person to learn what they are doing. Two other subjects, although recognizing that the practice helped, voiced criticisms. Only one subject mentioned that this does sometimes become a hindrance "*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? So in this case I kept [the junior analysts] on for a long time but at some point when the entire team was very frustrated, I fired the rest of them. That way, the senior analyst could focus. ... You always need to be thinking about the next mission, not just about the mission you're currently on.*"

Another subject emphasized that for shoulder surfing to be effective, it has to be interactive. "*Shoulder surfing isn't always the best. Just because if the new member isn't particularly motivated then they're just gonna be staring at a screen and not learning anything cause they don't want to and on the other hand, if the experienced person doesn't have the teaching mentality or doesn't really care to teach anybody, then they're also not gonna learn anything. They're just gonna be firing away on the keyboard. And the person shoulder surfing is just kind of like: I don't know what you just did, but it worked. <laugh>... it would be nice to kind of have more of a side saddle process where instead of me just watching you do it, let's both do it at the same time separately while you kind of show me how to do it. More of that instead of just watching it.*"

4.4.2 Task separation

Four subjects mentioned tasks being assigned by team leads to analysts. Two of these subjects discussed that their teams did this task assignment based on the skill level or experience of the analysts. One of these subjects' responses was especially interesting because they used the Pyramid of Pain (see §4.1.3.3) to assign tasking: "*We try to siphon things towards the top of the Pyramid of Pain to our master analysts and let our analysts on-site deal with the middle part.*" A third subject mentioned that tasks ought to be assigned by the skill level of the analysts but this was not always the case. They discussed: "*I guess that's*

just a matter of the [team lead] understanding which tasks can be done by less experienced [analysts] and when we're off mission focusing more on dummy proofing the [documentation] for those specific tasks."

The last subject mentioned that because of task separation, analyst note-taking and adhering strictly to the process becomes important so that every member of the team can stay on the same page. "*I would say those steps are followed pretty strictly because ... if you're dividing ... work between people ... It's easier to say: 'you focus on host based indicators, make sure uploading [notes] here.' And that kind of helps the flow of the team. ... If people are kind of injecting themselves ... in the middle of a step, it kind of throws things out of whack and makes it hard to kind of track a mission from execution to completion.*" Good notes and process adherence may help facilitate better analyst tracking. This seems related to the discussion of analysts going down rabbit holes in §4.2.5. These behaviors also seems related to certain personalities facilitating teamwork better than others (see §4.5.5) as certain personalities may naturally document their activity better or follow tasking more rigidly than others.

Many subjects volunteered that certain tasks fit newer members better than other tasks. Three subjects mentioned that new members were important for providing feedback from an outsider's perspective (see §4.3.1). Two subjects discussed newer members being capable of discrete, straightforward tasks compared to more vague tasks that should be given to experts. Two examples that were given were looking at the top 10 most common DNS requests made from a DNS server or "*proving negative things*" by going line by line through a log file to make sure nothing out of the ordinary is in it. Different individual subjects also suggested: being an extra set of "*eyes on target*" to catch "*low hanging fruit*"; looking up information like "*version type and what vulnerabilities [are] related to that version*" on different customer applications; and helping other new members get up to speed because they know where all the documentation is. Failure to divide up tasks could result in analysts not being used effectively. One team lead described a recent exercise their team was deployed on: "*We found half our people to be sitting around, doing nothing, most of the time. Cuz I can't tell somebody to go pull a memory dump and do some memory forensics and they're like: I don't know what that is.*"

4.4.3 Process Documentation

Six subjects described good process documentation as important for assisting new members. Two reasons were given for this. Two subjects brought up that good documentation was required to ensure a minimum standard was being met. One said: "*If you have a team that you don't know well, or they're newly coming on, or there's a lot of turnover, then I would say more emphasis needs to be in the checklist because you have to make sure where you're getting a minimum quality standard*". Four other subjects discussed how good documentation would provide guidance for newer members. They made comments similar to: "*It helped them focus on each task at a time.*" and "*I think having the [documentation] will help them be more effective, faster, because it will give them a guideline of what to do. ... 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.'*".

Detailed process documentation also helps members with less cyber expertise. Five subjects asked for more documentation in §4.2.3, of whom four had less than two years of TH experience. The only experienced subject in that group was asked if the documentation should be more or less vague than it currently is and they responded: "*I would say more specific, especially for a new team like us*". As mentioned in §4.2.3 the only subject that thought the current process documentation was too specific, later indicated more documentation would be helpful for newer members. This subject had 2 years of experience. One subject also discussed that newer members follow processes more closely than more experienced members: "*Usually the less experienced ones are better at following the process, mostly because they don't know any better to do anything else.*"

4.4.4 Other

Even though the subjects were being asked specifically about process changes that could be made to assist members on a mission, some subjects brought up suggestions that dealt with training and personnel issues. I report these here to be consistent with the protocol.

4.4.4.1 Training

Three subjects suggested training was important, with two subjects mentioning on-the-job training specifically: "*I think on-the-job training is really good.*" and "*we've got a process. [Describes on-the-job training with different teams] I think what's really important is to have some kind of pipeline like that.*" Two of the three subjects seemed primarily concerned with getting newer members time with the tools. When asked for hypothetical process changes they said: "*More of those range type environments — I would love that. We don't have that right now, we are working on it, but where I can just have the team log on in their free time and just play around, use it, and then the tools are already there for them to use. ... they can just play around with the tools and get comfortable with it.*" These results agree with the findings in §4.5.2.

4.4.4.2 Personnel Issues

In addition to the subject above that thought a training "*pipeline*" was important, one other subject believed recruiting newer members with a better background in cyber-related fields would help. "*I think one big variable will be having [a cyber specialty within the organization] so that way, they report to their team with their pipeline training complete*".

4.4.5 Counterproductive Processes

There was little consensus on what processes were counterproductive to the integration of newer members. Two subjects discussed that they did not have enough time to provide their newer members with guidance while on site. One subject said: "*If anything, it was definitely the time constraint and we're making a change for the better where we can be there for as long as we need based on the size of the network. ... In my experience analysts only have so much time to hopefully get work done so they can't spend too much time answering questions and just doing show and tell kind of with the [junior] analyst.*" The other subject also talking about teaching newer members while on mission said: "*You have a situation where it would be nice to spend the next two hours teaching 'em how to do this tool or whatever it is, but*

you know, we've already done the minimum of what we need to do to move on to the next step and we kind of just have to [say], okay, now next tool, next tool, next tool. And they don't get the time to sit down and really dig into and get experience with one thing. And then the next thing they're kind of just getting this real quick and dirty. Here's what this tool does. Okay. Now the next tool, here's what this tool does and we don't really have the time to get them to sit down and, and really learn.".

Change in general was mentioned by three subjects as being counterproductive. Two subjects were concerned about too much change in the process, especially before newer members had the opportunity to see why the change was being made. A third subject mentioned a change in the toolset as also being counterproductive. One subject mentioned ambiguity in tasks and objectives. This seems directly related to the discussions about mission planning in §4.1.3.2. Lastly, although detailed documentation was seen as helpful (see §4.4.3), one subject "*we could probably, for the lack of better words, dumb it down a little bit. The guy who, who created the playbook was pretty smart and technical. Um, so we could probably dumb it down a little bit and still have it work.*"

4.5 RQ2.3: What features indicate expertise to Threat Hunt team members?

In this section, I report on the features that subjects identified as indicating expertise among threat hunters. In this section, not every subject weighs in on every feature. In the first subject's interview, no features were specifically asked about. In the following interviews, features that had been mentioned unprompted by previous subjects were brought up in a follow-up question after asking the subject what features indicate expertise and allowing them to respond. This ensured we got many views prior to biasing subjects by asking about specific features. Table 4.5 indicates the most discussed features and how many subjects viewed them positively or negatively. Each feature is elaborated below.

4.5.1 Experience and number of missions

11 subjects explicitly mentioned experience as a feature or discussed the number of missions an analyst had been on. Six were code as viewing experience as a feature that indicates

Table 4.5. Subjects explicitly reported or contradicted these features as indicative of TH expertise or potential.

Features that indicate expertise or potential	**# subjects (from # orgs) spoke positively about feature**	**# subjects (from # orgs) spoke negatively about feature**
Experience and number of missions (§4.5.1)	6 (3)	5 (2)
Training (§4.5.2)	3 (2)	4 (2)
Certifications (§4.5.3)	2 (1)	7 (3)
Doing cybersecurity work in personal time (§4.5.4)	7 (2)	1 (1)
Personality (§4.5.5)	5 (2)	2 (1)
Curiosity (§4.5.6)	3 (3)	0 (0)

expertise. For example, when subjects were asked about what indicates expertise both these responses: "*I would say for several [analysts], that number of missions is a good quantifier.*" and "*Just their, their hands-on experience, I guess.*" were coded as subjects that believe experience is an important component of expertise.

However five subjects disagreed. The remaining subject was not clear on their position so they were not coded into either group. The dissenting subjects pointed out cases where they knew analysts that had been on many missions or been threat-hunting for a long time and yet were not perceived as experts. One subject said: "*Experience does not equal quality ... just because someone has experience doesn't mean they're good at their job.*" This point was also conceded by many that held that expertise was important. For example, the subject indicating a correlation between success and number of missions also said: "*You also have people who've been on a hundred missions and they just don't have the aptitude or the thought ability. You also 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. I would say there's very little correlation between ... [lists many things that do not contribute to expertise. These are discussed later]*" This is important because, in addition to the results being divided, those that affirmed experience as an indicator of expertise often did so hesitantly and recognized that it was a general rule rather than a predictor in every case.

4.5.2 Training

Only one subject spoke favorably of training as an indicator of expertise. The person was an inexperienced member of leadership. They said: "*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. I don't spend as much time in the tools as I probably should so I don't, it's hard for me to tell an analyst what he should be looking for or what they should be looking for.*" Two more experienced subjects seemed more cautious saying "*there is a mild or weak correlation between number of training*

courses and analyst success" and "*it depends on the training they've had. ... I feel like it has to do with training and your experience level. And your, what's the word, excitement about your job, I guess.*". Four subjects spoke unfavorably of training as an indicator of expertise. One experienced subject said: "*I don't know. It's hard. We've been trying to do training, ... We do weekly brown bags/workshops. We try to set up time with them to give them more exposure to our kits, to our tools. Our master analysts and contractors, they'll give them very hands-on training, overview of stuff. And to be honest with you, it's hard to notice any difference. Like maybe long term it will create an impact, but it doesn't change it immediately, I think.*" An inexperienced team lead said "*having a [certain] course under your belt or any other course, usually doesn't give me immediate confidence.*". One analyst felt very strongly about certain training courses saying "*some of these trainings we have [an agency] requirement. I'm not sure if you ever took [name of course], but I consider it just a massive waste of time. <laugh>*" However, the analysts recognized that while training did not indicate expertise, training was helpful for bringing people up to speed more quickly: "*we have done one-off trainings. I've given the team several trainings just about [tool] and I have done like stuff on the assess kit but we have done extra stuff like that and just little exercises to help people get up to speed a little quicker.*" Another subject agreed that training could reduce time to integration for newer members. When asked how to reduce that time they said: "*Obviously training is a huge one. Specifically I would say like the two biggest types of training are forensics ... either like host forensics or network forensics ... and then just tool specific training. That's a huge part of it, is just like being comfortable with how the tool works.*"

The importance of time with the tools was a theme that often accompanied discussions about training whether as an indicator of expertise or as a way of integrating newer members. The value of training was reportedly correlated to the time spent on certain tools by four subjects. The first two were quoted above. The other two subjects spoke about two distinct training opportunities they had recently participated in. They are referred to as "training 1" and "training 2". The first subject said: "*looking at the [training 1] that we're going through, it's a good experience and it's a good training opportunity, but at the same time, 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." A second subject was asked: "*Thinking back to [training 2]. Was there a time during [training 2] where you felt like the process was hindering your ability to actually like do an effective hunt?*"

Subject: "*Yes and no. At a high level, I enjoyed the process. It kind of gave us a way forward but exerc-isms aside — all the range-isms and all that stuff — really, it was more of an inability to adapt it specifically to the toolset we were given. Just because we couldn't bring our own kit, it was all their own hosted stuff. I don't know that a change in process would've helped us any more other than again having more collective granular data stores or tools that we could just pull and bring with us.*"

Interviewer: "*Is that a common experience when you're not able to use your own kit, do you think?*"

Subject: "*Yeah. Pretty much any time you go to an exercise like that and we have to use their stuff. It's, you're spending like the first day figuring out what the heck to do, [because] you don't have your own stuff, you can't plug in and go*".

4.5.3 Certifications

While some certifications are tied to a specific training, most certifications can be earned by passing a test. Since many cybersecurity certifications are not necessarily tied to any specific training [91], [92], they were included as a different category.

Seven of the subjects did not believe certifications or "certs" were a good measure of expertise. For example, they said: "*I would say there's very little correlation between number of certifications and analyst success.*", "*I don't put personally a lot of weight into certifications because I think a lot of times those are very test based and I don't know that they have a lot of application with the real world.*" , and "*I think some, but not really. I know some people who are very well certified, [but] they're not as much of an expert*". Two additional subjects said certifications were only good for proving that a hunt team member has a baseline amount of knowledge. One of these subjects, when asked a question unrelated to expertise said: "*the certs are good and we all do [organization] courses and stuff like that. I think those are very helpful, great starting ground but can't expect people to get those certs and just be good and*

be ready to go out on a hunt mission. They definitely get the certs as an entry point and then start working with the tools to understand them." and then when asked if certification indicates expertise they said: "*I've seen several people that have, you know, plenty of certs and, you know, they're not, I don't think they're as proficient as the certs says they are.*". The other subject speaking in the response to a question about expertise said: "*Certifications, I think are a good baseline, but not what should necessarily be used beyond that. It's good to know people have, you know, your [name of cert] or [name of cert] or at least some sort of "coming in baseline" where they understand a little bit, but experience actually being on a network, threat hunting [is] more valuable once you've gotten that baseline.*" One subject that did not believe certifications indicated expertise also mentioned specifically that they did not think certifications were a good measurement even of baseline knowledge: "*It's really hard to define what a baseline of cybersecurity understanding is, right? ... 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.*" None of the remaining three subjects discussed certifications in reference to expertise. No subject specifically mentioned that they thought certifications were a good measure of expertise.

One organization had a specific internal certification document that was mentioned by name by five subjects. One subject was stating that by using this document, the team is able to be interoperable with teams from other governmental organizations. Another subject simply referenced the document in passing. The other three subjects voiced dissatisfaction with the document as it relates to indicating threat hunt member expertise. One subject thought it was too trivial and not relevant: "*we have the [document], which, I don't know, I have some mixed feelings on the [document]. ... It doesn't really apply to us all that well and the proficiency needed to complete it isn't really that much either.*". Another subject concurred with the relevancy point, wishing it was more practical saying: "*my own personal perspective on [the document] is in the [organization] and for most jobs related to cybersecurity, is we need to get rid of them, ... Why not have someone use [training tool] and do actual incident response. ... I wouldn't say that the [document] actually adds anything to their knowledge. I think most of their knowledge comes from what they've done outside of that specific document. It's funny, we were having this conversation yesterday with some of*

the [team] leads as far as how we make something better." The last subject didn't think the internal certification could be used to measure expertise because it was not being followed sufficiently rigidly.

Interviewer: "*If you had to have metrics to determine whether someone was an expert or not, what metrics would you use to try to describe that?*"

Subject: "*So our metrics that we use right now are our [document] ... so that's what we use now.*"

Interviewer: "*Do you personally find that to be a good measure of expert versus non expert analysts?*"

Subject: "*Right now, I don't think it's that great of a measurement because we have to wave a lot of things because we just haven't been [a team] long enough to have people that have been in their role for five years or done two [different jobs]. I think eventually it can be, but right now we're so new that we're still kind of filling in the patches.*"

4.5.4 Off-The-Clock Work

Ten subjects discussed the significance of TH team members working on their own time. Seven subjects said doing cyber security activities on personal time was a good metric for potential or expertise. Four of these subjects said it was a good metric of potential, two said expertise, and one said both. Subjects that viewed off-the-clock positively made comments like: "*Maybe this is more like a personal thing, but sometimes I like to see or hear, if they're doing it on their own time. If they're doing like any kind of CTFs or something on their own time or researching stuff. ... Even if they may not be super proficient, they're at least willing to learn and put some extra time into it. ... That's definitely more of a potential thing.*", or "*I can't really put my finger on a single thing that kind of 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 timeframe, they have a lot of free time where they can kind of play around with the tools or do you know, capture the flags, the different challenges that come throughout the year that almost has been a bigger indicator.*", or "*Yeah, definitely.*

... We don't really expect them to go home and you know, until midnight they're working on [tool] scripts, but some people do, some people just enjoy that ... but definitely if somebody's, somebody's putting their free time into it, you can expect that they're gonna perform a lot better."

Three subjects mention Capture The Flags or CTFs. They were not included as their own theme in this section as the three subjects seemed to have different ideas of what they indicate. The two subjects above mentioned CTFs as something done in the members free time while the third subject discusses CTFs in the context of personality (this subject is quoted in §4.5.5).

One subject didn't think working during off hours was necessarily associated with either expertise or potential. Instead, they correlated it with passion and the speed of advancement. They said: "*I think it's a good metric for passion. ... I think passion means you can go faster or advance faster. ... The person that's just sitting on their computer, researching networks all night might get there faster than somebody that does it on clock.*". Another subject did not mention working off hours specifically but was concerned about the mental well being of TH team members who work too much in the long term saying: "*I don't know that I'm qualified to say this since I feel new myself but ... I feel like we've got a very work-hard-play-hard culture where we get our work done, but we also like to have a lot of fun. I think someone's going to commit their entire life to doing this is gonna get burned out really fast.*" Burnout is the result of job related stressors effecting a person over a long period of time [93]. One of the subjects in leadership was also concerned about burnout and other issues that could result in working off hours but they did believe that working off hours was indicative of potential in a TH member. This subject implemented a rule, not allowing their personnel to work more than a certain number of hours per day. When asked if working during off hours and one other metric were measures of expertise, potential, or something else, this subject responded: "*I think they definitely are good indicators of potential. I know one thing I'll mention in this context is that we have on our team, we've set limits for operators on how much they can do analysis on [a mission], because what we didn't wanna have happened was have, essentially burnout where you know, some of our things are remotely accessible, not only on a client location so what we didn't wanna have happen was people going back*

home or going back to a hotel or going somewhere else and VPNing back into, you know, the environment and then starting to, you know, spend all night hunting but without their teammates so we did set some limits on: you're only going to look at, you're only gonna be on [a mission] for, I can't remember the specific times, I think we put a certain week limit and a certain day limit."

One other subject also mentioned working on personal time as a positive in the context of learning new material quickly. "*Also just a drive, for kind of learning. One of our best analysts really had no background in cyber security [or] information security, but was so willing to put the time in to learn that now they're thriving, but that's hard when it comes to finding that specific person.*"

4.5.5 Personality

Five subjects indicated that personality was a good indicator of expertise or potential. These subjects make statements like: "*You ... have people who've been on a hundred missions and they just don't have the aptitude or the thought ability. You also 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. So some of it's personality driven, some of it's training driven.*", and "*So just because someone has experience doesn't mean they're good at their job. But I would say it comes from the personality. Some people are passionate about this stuff, they go above and beyond on their own. They do lots of CTFs, and training. So it depends on the personality.*" Two subjects directly connected personality, not to expertise but instead to teamwork. One said: "*Personality factors into team dynamic, but not necessarily expertise.*" and the other said: "*I don't know about directly. I think that more feeds into how the team can progress when you have certain personalities. Sometimes you'll run into, if you have a more experienced member who really doesn't like talking about what they're doing, it just wants to sit in a corner and, and do stuff that can kind of hinder your forward movement versus somebody who's really outgoing about what they're finding.*" Two additional subjects indirectly connected personality to teamwork. The first additional subject tied certain jobs in a TH team with different personalities, saying: "*Somewhat? Yeah. I mean, they have to have*

that personality, that, to, that they want to do this, or I guess that's not a personality trait, but the desired, um, <trails off><laugh>That's kind of an interesting question. Just because in dealing with customers, you would want somebody that has the personality that could mingle with other people. Typically cyber people aren't that outgoing or not outgoing, but they're a little inclusive <laugh>". The second subject, when asked about personality, thought about the way a TH team operates: Either (1) coordinated as a team or (2) uncoordinated and effectively a group of individuals. They said at the end of the interview: "*No, I guess like one thing, the personality question made me kind of think of this, but another thing I see a lot of is the hunting as a team is, or any like cybersecurity evolution as a team is, no more effective or even less effective than individuals doing it on their own because there's no like proper coordination or like overall team, like there's no division of work or there's no, like there's not, or there's not like smart division of work and smart goals that the team's being pointed towards.*"

4.5.6 Curiosity

Curiosity was mentioned unprompted by two subjects as something indicating potential. They made statements like "*I think someone who is either very creative or very investigative, like very interested in learning about something very deeply, from an academic perspective so intellectually curious.*" or "*I think as long as you're curious and you're willing to keep researching everything you're finding and learning new things on the fly ... a really good threat hunter will be able to do that quickly and determine whether or not something is good or bad using all available resources versus just saying: this is online so it's good, right?*". Another subject said it was the biggest indicator of expertise. "*I would say the number one trait in my per perspective is curiosity and ability to abstract. So if I'm looking for an analyst, I want somebody who's very curious. Who's willing to chase down a rabbit hole even if they think it's a waste of time, [because] they wanna learn. And if they're someone who does that, then they end up finding more things than not*". Additionally, two other subjects gave examples of members asking many questions as indicative of potential. One said "*usually you can tell as soon as you either meet them or the questions they're asking*

about the process. I've had a guy ... he's already sent me and [coworker] an email about like, what's the qualification process? How do I get qualified as fast as possible? ... What do I need to get started? And usually you can tell just by the questions they're asking versus you having to tell them Hey, this is how you get qualified. Here's your deadline. There's usually a pretty big divide between those people and how quickly they pick up the material and get on keyboard." The other said: "*No, I remember, when I was there, when [an expert analysts] reported in, and I could tell he was going to be a good team player, like a good technical analyst, because he started asking millions of questions. He was talking to people, he was trying to learn stuff. I would say, right off the bat, they're trying to be part of a team and learn. But then for non-achiever, it's also noticeable. People who just kind of are not proactive enough, they'll sit there and wait ... so you can tell, right from the start you can tell if the person's going to be a good team member and good technical analyst, versus someone who does not want to be there.*"

5. DISCUSSION & FUTURE WORK

5.1 Recommendations for TH Teams

Having a process assists with threat hunting [69], particularly in the presence of high turnover [18]. I make six recommendations for TH processes based on the recommendations, emphases, and inconsistencies observed in the interviews.

5.1.1 Mission Planning

Given the importance of objectives in subject interviews, (discussed in §4.1.3.2) creating and distributing the objectives to the team needs to be a priority in any TH process. One of the subjects in leadership said: "*the critical driving piece is that mission planning, scoping that happens prior to actually deploying agents in prior to people actually getting on site. I think that's actually the most critical piece of all of this ... Cause I think those are the things we've learned from having bad engage — not bad engagements, but not, not real worthwhile engagements. This section of the planning process is really what I think creates a useful engagement.*"

The creation of a hypothesis or similar planning document should be required and the objectives should be shared with everyone on the team. Analysts might want to create their own hypotheses, especially if they help analysts maintain focus as one subject claimed, but they should also have access to the higher level hypothesis as well. This method of creating multi-tiered hypotheses corresponds to what one subject recommended: "*so you plan the mission which is just dates, what your goals are, your high-level goals and then you'll push that into developing your specific tools and analytics.*" About half of the process in Figure 4.1 observed occurred before the team arrived onsite. This initial planning phase exists to ensure proper objectives are created and propagated through the team.

5.1.2 Balance Resources According to Probability of Success

As discussed in §4.1.6, APTs usually are sufficiently careful not to get caught by most alerting systems. If this is the case the amount of time spent on the automated alert loop

vs. the manual loop should be evaluated. The second subject quoted in that section said 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. Depending on the mission or the anticipated success of each loop in a given situation, the proportion of resources put into that loop should be regulated appropriately.

This imbalance between the detections made by the analysis loops could also be an issue with the alerts being used. One analyst was describing an adversary detection on a previous mission and was surprised that no alert was created for the activity. They said: "*I don't think there was an alert on it. Which is odd thinking about it now, because 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. If the rules are provided by the intelligence component then this issue could be a threat intelligence issue discussed later in §5.2.8. Regardless of the reason for the alerting gap, regulating resources to or from the Automated Alerting loop based on the quality and quantity of the indicators may assist with team efficiency.

5.1.3 Use TH Frameworks

Using predefined procedures is a sign of hunt team maturity [69] and using frameworks may aid consistency and effectiveness. Subjects indicated Mitre ATT&CK was the most used framework on their TH teams (see Table 4.1). Since it is popular among government TH teams, implementing this framework may increase interoperability with other teams. Mitre ATT&CK can also help provide examples and data sources for given TTPs which could mitigate a lack of documentation. This could be why it was so popular given that so many subject requested more documentation (see §4.2.3).

While it is unsurprising that Mitre ATT&CK was used as heavily as it was, many experts also recommend the use of Pyramid of Pain in cyber security [57], [94] and TH specifically [1]. Mitre ATT&CK is already geared towards items higher on the Pyramid of Pain but building the Pyramid of Pain into a process will also assist in ensuring that the team does the maximum amount of damage to the campaigns of any adversaries that are discovered on a

hunt mission. Only two subjects indicated they used the Pyramid of Pain and both qualified saying they used "*that concept*" or that it was not in the current standard but in the "*standard that we are creating*" (this subject was a member of leadership). Pyramid of Pain seems to lend itself especially well to government Threat Hunting as it models how to do the most damage to an adversary's offensive capability. Defending the US against such adversaries is one of CISA's explicit goals [95] and, to some extent, the goal of every government TH team.

In order to assist a TH team integrate the Pyramid of Pain, examples of where the Pyramid of Pain could fit into the process are shown in Figure 4.1. These task were selected because they require the analyst to prioritize some alerts, anomalies or IOCs over others. In accomplishing such tasks, prioritizing things higher on they pyramid of pain could be helpful.

5.1.4 Include Specific Process Documentation

As discussed in §4.2.3, subjects agreed that more process documentation would be helpful to newer members. SIEM queries were the most requested documentation addition so that might be a good place for teams to begin. The Mitre ATT&CK framework includes high-level indicators paired with data sources where they may appear [96]. Since many teams already use ATT&CK in their process, this might be a good starting place for the queries.

5.1.5 Perform Sensor Placement Before Deployment

When the team arrives on site they need data to begin hunting. To facilitate this, ensure that the team's sensors are deployed ahead of time. One subject mentioned work weeks being a natural network cycle. They said 2 or 3 weeks in advance of a mission makes the most sense.

5.1.6 Allow Training on Mission

The behavior of having new members "shoulder surf" with expert members is an apprenticeship activity but was not described in sufficient detail by any subject to verify that it fits into existing apprenticeship frameworks [97]. There is a large amount of literature on the

virtues and shortcomings of apprenticeship, especially formalized apprenticeship [98], [99]. Apprenticeship is being used with success in cybersecurity [100], [101]. A similar process, Pair Programming, is being used with success in software development where teams of two not only result in better learning, but also higher morale and better quality code [102]. Since subjects seem to agree that apprenticeship is the best way to integrate new members, teams may benefit from a more robust apprenticeship program. Since doing a hunt mission with trainees takes longer (§4.4.5), the team will need extra time on site. Considering the importance of on-the-job training to the subjects interviewed (see §4.4.4.1), this time is likely worthwhile. Subjects also mention that it is important to ensure trainees are able to interact with data, not just watch others hunt.

5.2 Future Work for Researchers

5.2.1 Develop Automation

Multiple suggestions for automation were made in §4.2.1. The most popular suggestions were automating the baselining task and automating the re-occurring kit set-up tasks. Most of the automation in TH literature focuses on automating the entire TH process [72]–[74], [78]. These automations being created may help bring balance to the issues described in §4.1.6 as they would fit into the automated analysis loop. However, these automations were not the types of automations that subjects requested. Automation that baselines or assists with kit set-up or does one of the other tasks remains for future work.

5.2.2 Open Questions

Multiple open questions are listed in Table 4.2. These are questions that my subjects seemed divided on indicating that there may not be a consensus among TH practitioners. Further research is needed to see if other TH teams have answered these questions satisfactorily. A survey among TH practitioners would be a good methodology for collecting this information.

5.2.3 Automation hindering analysts

In §4.2.1, one team lead believed automation would allow analysts not to have to think for themselves. A second subject in a leadership position indicated that Threat Hunting by definition is a manual process. Measuring if automations make equally capable teams less effective would be useful.

5.2.4 Automation Management

One team lead indicated that they were satisfied with the amount that was currently being automated but unsatisfied with how the automations were being managed. Software management has been a field of research for a long time [103]. Future research could investigate if this issue was a one-off issue with this team or a recurring issue within Threat Hunting due to unique Threat Hunting challenges.

5.2.5 Determine Which Data to Collect

One team provided an example of their team collecting too much data. Another subject said they were trying to change the mindset of collecting all possible data. This second subject indicated that precise objectives could help the issue of too much data. Even in this case, the scoping cannot be arbitrary and by scoping down the data being looked at, there could be adversary activity in other logs being missed by the team.

5.2.6 Process Definition vs. Flexibility

In §4.2.3, subjects request more detailed documentation in the TH process, especially for members with less expertise. However, in that section, an example is given of a mission where the process that was too precise, becoming a hindrance to the team. A second subject described some process documentation that was also too precise for expert TH team members, although may have been helpful for novice members. A good solution to this tension between process definition for novice members and flexibility for expert members is an open question for newer TH teams.

Some teams deal with this tension by allowing team leads to make changes to the process, however this only works for team leads with expertise. The problem is that not all organizations require their team leads to have TH experience so not all team leads will be capable of making these decisions. Between the three organizations being studied, the team lead group had by far the least cumulative experience. This may be a hazard for organizations that deal with the tension by allowing team leads to make changes. No other method for dealing with this tension was given by subjects.

5.2.7 Rabbit Holes

Rabbit holes can be useful if they lead to the discovery of an adversary but often they are simply a waste of time for the analyst. There is no way to predict if a rabbit hole will be worth going down until after it has been investigated. Having a way to make this determination ahead of time would be useful and help remove resentment between analysts who want to investigate things thoroughly and team leads who want to ensure coverage.

5.2.8 Cyber Threat Intelligence

Dissatisfaction with a team's intelligence component was a common complaint (see §4.2.4). Since intelligence is so important according to the Threat Hunting maturity model [69] and other processes [1], this could indicate a serious issue. Recommendations were provided regarding indicators since this seemed to be one of the most common ways a team and its intel component interacted. In addition to these recommendations, researchers are working on defining what quality CTI is [104], [105] and how to best share it with operators [106]. It was not clear in interviews if sub-par indicators was subjects' only complaint. More research investigating the relationship between a TH team and its intelligence component, while considering expectations could help clarify this.

5.2.9 Process Evaluation

The goal of Threat Hunting is to reduce adversary dwell time [1] but this metric cannot always be used. Since government TH teams do not necessarily consistently hunt on the

same networks, measuring dwell times over time is not possible. Some organizations also may be compromised so infrequently that measuring dwell times does not provide a Hunt Team with actionable feedback.

Finding a metric to replace dwell times is an open issue. Most metrics cannot account for a TH team overlooking an adversary because if an adversary gets overlooked by a TH team, the organization may never know. Many metrics have been suggested to measure how useful TH missions are [1], [58] but none are as objective as dwell time. Instead, these metrics measure if the TH mission added value to the organization by pointing out vulnerabilities or security blind spots. This issue of objective measurement is similarly difficult across other cybersecurity fields [107]. An objective way to measure the effectiveness of a TH team may help evaluate how good a process is.

6. SUMMARY

84% of organizations perform threat hunting. Over half of these organizations lack a formal methodology, and instead perform their threat hunting *ad hoc* [8]. In this work, I provide the first government TH process model (see Figure 4.1) created from subject interviews and drawn process diagrams. This process can be used as a starting point for organizations without a process or for comparison for organizations with processes. In addition to the TH process, I describe process recommendations provided by subjects. I enumerate factors described by subjects as expediting new member integration, recommendations for faster integration and features that indicate expertise to TH team members. Much work remains for future researchers. Many open questions have been presented including how to ensure automation does not hinder TH analysts and how to determine what data to collect. The relationship between a TH team and its intelligence component was another common difficulty brought up by subjects where future work may be helpful. In the long term, measuring the effectiveness of TH processes is an open problem, which is further complicated by the sensitivity of doing experiments with a Threat Hunting team.

REFERENCES

[1] R. van Os, M. Bakker, R. Bouman, M. D. van Leeuwen, M. van der Kraan, and W. Mentges, *TaHiTI: A threat hunting methodology*, Dec. 2018. [Online]. Available: https://www.betaalvereniging.nl/wp-content/uploads/TaHiTI-Threat-Hunting-Methodology-whitepaper.pdf.

[2] "The Cost of Cybercrime," Ponemon Institute LLC and Accenture, Tech. Rep., 2019.

[3] *National Cyber Strategy*, Sep. 2018. [Online]. Available: https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf.

[4] D. E. Brink, "Quantifying the Value of Time in Cyber-Threat Detection and Response," en, Aberdeen Group, Tech. Rep. 15218, Jan. 2017.

[5] "Cost of a Data Breach Report 2020," en, IBM Security, Tech. Rep., Jul. 2020. [Online]. Available: https://www.ibm.com/security/digital-assets/cost-data-breach-report/1Cost%20of%20a%20Data%20Breach%20Report%202020.pdf.

[6] W. R. M. Lee and R. Lee, *The Who, What, Where, When, Why and How of Effective Threat Hunting*, en, Feb. 2016.

[7] *National Cyber Strategy*, Mar. 2023. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf.

[8] W. R. Lee and R. M. Lee, *The Hunter Strikes Back: The SANS 2017 Threat Hunting Survey*, en, 2017.

[9] K. Wafula and Y. Wang, "CARVE: A Scientific Method-Based Threat Hunting Hypothesis Development Model," in *2019 IEEE International Conference on Electro Information Technology (EIT)*, ISSN: 2154-0373, May 2019, pp. 1–6. DOI: 10.1109/EIT.2019.8833792.

[10] A. Agarwal, H. Walia, and H. Gupta, "Cyber Security Model for Threat Hunting," in *2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Sep. 2021, pp. 1–8. DOI: 10.1109/ICRITO51393.2021.9596199.

[11] J. R. Bynum, "Cyber Threat Hunting," Master's thesis, Utica College, 2019.

[12] K. McIver, "Closing the Revolving Door," Tech. Rep., Aug. 2022. [Online]. Available: https://icitech.org/wp-content/uploads/2022/08/Closing-the-Revolving-Door.pdf.

[13] J. Oltsik, "The Life and Times of Cybersecurity Professionals 2018," en, *Research Report*, 2019. [Online]. Available: https://www.esg-global.com/hubfs/pdf/ESG-ISSA-Research-Report-Life-of-Cybersecurity-Professionals-Apr-2019.pdf.

[14] *How High Employee Turnover Poses Increased Cyber Security Risk*, en-US, Nov. 2020. [Online]. Available: https://copperbandtech.com/high-employee-turnover/.

[15] *Cyber Security Analyst Demographics And Statistics In The US*. [Online]. Available: https://www.zippia.com/cyber-security-analyst-jobs/demographics/.

[16] A. Ju, H. Sajnani, S. Kelly, and K. Herzig, "A Case Study of Onboarding in Software Teams: Tasks and Strategies," in *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, ISSN: 1558-1225, May 2021, pp. 613–623. DOI: 10.1109/ICSE43902.2021.00063.

[17] P. Rodeghero, T. Zimmermann, B. Houck, and D. Ford, "Please Turn Your Cameras On: Remote Onboarding of Software Developers during a Pandemic," *arXiv: 2011.08130 [cs]*, Mar. 2021, arXiv: 2011.08130. [Online]. Available: http://arxiv.org/abs/2011.08130.

[18] T. Nosco, J. Ziegler, and Z. Clark, "The Industrial Age of Hacking," en, in *Proceedings of the 29th USENIX Security Symposium*, Aug. 2020.

[19] A. White and B. Clark, *BTFM: blue team field manual*, eng, Version 1, Rel 2. United States: Alan White, 2017, ISBN: 978-1-5410-1636-1.

[20] C. Donalds and K.-M. Osei-Bryson, "Cybersecurity compliance behavior: Exploring the influences of individual decision style and other antecedents," en, *International Journal of Information Management*, vol. 51, p. 102 056, Apr. 2020, ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2019.102056. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0268401218312544.

[21] N. Alharbi, "A Security Operation Center Maturity Model (SOC-MM) in the Context of Newly Emerging Cyber Threats," English, ISBN: 9798672151229, Ph.D. The Claremont Graduate University, United States – California. [Online]. Available: https://www.proquest.com/docview/2447834861/abstract/49F3D76000004605PQ/1.

[22] S. Y. Cho, J. Happa, and S. Creese, "Capturing Tacit Knowledge in Security Operation Centers," en, *IEEE Access*, vol. 8, pp. 42 021–42 041, 2020, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2976076. [Online]. Available: https://ieeexplore.ieee.org/document/9007685/.

[23] S. C. Sundaramurthy, J. Case, T. Truong, L. Zomlot, and M. Hoffmann, "A Tale of Three Security Operation Centers," en, in *Proceedings of the 2014 ACM Workshop on Security Information Workers - SIW '14*, Scottsdale, Arizona, USA: ACM Press, 2014, pp. 43–50, ISBN: 978-1-4503-3152-4. DOI: 10.1145/2663887.2663904. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2663887.2663904.

[24] P. Jacobs, A. Arnab, and B. Irwin, "Classification of Security Operation Centers," en, in *2013 Information Security for South Africa*, Johannesburg, South Africa: IEEE, Aug. 2013, pp. 1–7, ISBN: 978-1-4799-0808-0. DOI: 10.1109/ISSA.2013.6641054. [Online]. Available: http://ieeexplore.ieee.org/document/6641054/.

[25] M. Majid and K. Ariffi, "Success Factors for Cyber Security Operation Center (SOC) Establishment," en, 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: EAI, 2019, ISBN: 978-1-63190-198-0. DOI: 10.4108/eai.18-7-2019.2287841. [Online]. Available: http://eudl.eu/doi/10.4108/eai.18-7-2019.2287841.

[26] D. Sacher-Boldewin and E. Leverett, "The Intelligent Process Lifecycle of Active Cyber Defenders," en, *Digital Threats: Research and Practice*, vol. 3, no. 3, pp. 1–17, Sep. 2022, ISSN: 2692-1626, 2576-5337. DOI: 10.1145/3499427. [Online]. Available: https://dl.acm.org/doi/10.1145/3499427.

[27] M. Nyre-Yu, R. S. Gutzwiller, and B. S. Caldwell, "Observing Cyber Security Incident Response: Qualitative Themes From Field Research," en, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 63, no. 1, pp. 437–441, Nov. 2019, ISSN: 2169-5067, 1071-1813. DOI: 10.1177/1071181319631016. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1071181319631016.

[28] S. J. Roberts, *Incident Response is Dead Long Live Incident Response*, en, Section: posts, Apr. 2015. [Online]. Available: https://sroberts.io/posts/incident-response-is-dead-long-live-incident-response/.

[29] P. Cichonski, T. Millar, T. Grance, and K. Scarfone, "Computer Security Incident Handling Guide : Recommendations of the National Institute of Standards and Technology," en, National Institute of Standards and Technology, Tech. Rep. NIST SP 800-61r2, Aug. 2012, NIST SP 800–61r2. DOI: 10.6028/NIST.SP.800-61r2. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-61r2.pdf.

[30] S. C. Sundaramurthy, J. McHugh, X. S. Ou, S. R. Rajagopalan, and M. Wesch, "An Anthropological Approach to Studying CSIRTs," *IEEE Security Privacy*, vol. 12, no. 5, pp. 52–60, Sep. 2014, Conference Name: IEEE Security Privacy, ISSN: 1558-4046. DOI: 10.1109/MSP.2014.84.

[31] A. Naseer, H. Naseer, A. Ahmad, S. B. Maynard, and A. Masood Siddiqui, "Real-time analytics, incident response process agility and enterprise cybersecurity performance: A contingent resource-based analysis," en, *International Journal of Information Management*, vol. 59, p. 102 334, Aug. 2021, ISSN: 02684012. DOI: 10.1016/j.ijinfomgt.2021.102334. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S026840122100027X.

[32] R. Werlinger, K. Muldner, K. Hawkey, and K. Beznosov, "Preparation, detection, and analysis: The diagnostic work of IT security incident response," en, *Information Management & Computer Security*, vol. 18, no. 1, S. M. Furnell, Ed., pp. 26–42, Mar. 2010, ISSN: 0968-5227. DOI: 10.1108/09685221011035241. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/09685221011035241/full/html.

[33] D. Everson and L. Cheng, "Network Attack Surface Simplification for Red and Blue Teams," in *2020 IEEE Secure Development (SecDev)*, Sep. 2020, pp. 74–80. DOI: 10.1109/SecDev45635.2020.00027.

[34] M. N. S. Miazi, M. M. A. Pritom, M. Shehab, B. Chu, and J. Wei, "The Design of Cyber Threat Hunting Games: A Case Study," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*, Jul. 2017, pp. 1–6. DOI: 10.1109/ICCCN.2017.8038527.

[35] *Cybersecurity Experts Hunting for Hackers*, en. [Online]. Available: https://www.nationaldefensemagazine.org/articles/2017/4/3/cybersecurity-experts-hunting-for-hackers.

[36] H. Rasheed, A. Hadi, and M. Khader, "Threat Hunting Using GRR Rapid Response," in *2017 International Conference on New Trends in Computing Sciences (ICTCS)*, Oct. 2017, pp. 155–160. DOI: 10.1109/ICTCS.2017.22.

[37] D. Milea, *Hypothesis in Threat Hunting*, en, Jul. 2017. [Online]. Available: https://medium.com/@demetriom/hypothesis-in-threat-hunting-4bea5446e34c.

[38] P. Gao, F. Shao, X. Liu, X. Xiao, Z. Qin, F. Xu, P. Mittal, S. R. Kulkarni, and D. Song, "Enabling Efficient Cyber Threat Hunting With Cyber Threat Intelligence," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, ISSN: 2375-026X, Apr. 2021, pp. 193–204. DOI: 10.1109/ICDE51399.2021.00024.

[39] *Executive Order 14028: Improving the Nations Cybersecurity*, en-US, May 2021. [Online]. Available: https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/.

[40] *The United States Government Manual*. [Online]. Available: https://www.govinfo.gov/content/pkg/GOVMAN-2022-12-31/xml/GOVMAN-2022-12-31.xml.

[41] *Threat Hunting*, en. [Online]. Available: https://www.boozallen.com/expertise/cybersecurity/threat-hunting.html.

[42] *Falcon Overwatch: Managed & Proactive Threat Hunting CrowdStrike*, en. [Online]. Available: https://www.crowdstrike.com/endpoint-security-products/falcon-overwatch-threat-hunting/.

[43] *Talos Incident Response*. [Online]. Available: https://talosintelligence.com/incident_response/hunting.

[44] *6 U.S.C. ğ659*.

[45] R. Symonds, "Innovating the Prioritization of Cyber Defense," *Journal of Information Warfare*, vol. 16, no. 2, pp. 12–18, 2017, Publisher: Peregrine Technical Solutions, ISSN: 1445-3312. [Online]. Available: http://www.jstor.org/stable/26502753.

[46] S. M. McClanahan, *169th Cyber Protection Team Highly Capable, Always Ready*, en-US, Nov. 2019. [Online]. Available: https://news.maryland.gov/ng/2019/11/06/169th-cyber-protection-team-highly-capable-always-ready/.

[47] *CYBER 101: Hunt Forward Operations*, en-US. [Online]. Available: 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.

[48] "(U) Combat Mission Teams and Cyber Protection Teams Lacked Adequate Capabilities and Facilities to Perform Missions," Inspector general of the US Department of Defense, Tech. Rep. DODIG-2016-026, Nov. 2015. [Online]. Available: https://media.defense.gov/2017/Sep/15/2001810710/-1/-1/1/DODIG-2016-026%20(REDACTED).PDF.

[49] M. F. A. Network, *Effects of Moving on Military Families*, en-US. [Online]. Available: https://www.mfan.org/topic/moving-permanent-change-of-station/effects-of-moving-on-military-families/.

[50] E. Bledsoe, *How Often Do Military Families Move? Why They Move So Much?* en-US, Jan. 2023. [Online]. Available: https://www.thesoldiersproject.org/how-often-do-military-families-move/.

[51] E. Schmid, *Military families often have to move every few years. Critics say it's disruptive and unnecessary*, en, Section: Military, Apr. 2022. [Online]. Available: https://wusfnews.wusf.usf.edu/military/2022-04-03/military-families-often-have-to-move-every-few-years-critics-say-its-disruptive-and-unnecessary.

[52] U. S. G. A. Office, "Coast Guard Workforce Planning Actions Needed to Address Growing Cyberspace Mission Demands," Report to Congressional Committees GAO-22-105208, Sep. 2022. [Online]. Available: https://www.gao.gov/assets/gao-22-105208.pdf.

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

[54] Applebaum, Andy, Nickels, Katie, Schulz, Tim, Strom, Blake, Wunder, John, and A. A. K. N. T. S. B. S. J. Wunder, *Getting Started with ATT&CK*, en, A. Pennington, Ed., Oct. 2019. [Online]. Available: https://www.mitre.org/sites/default/files/2021-11/getting-started-with-attack-october-2019.pdf.

[55] S. Team, *A Framework for Cyber Threat Hunting Part 1: The Pyramid of Pain*. [Online]. Available: https://www.threathunting.net/files/A%20Framework%20for%20Cyber%20Threat%20Hunting%20Part%201_%20The%20Pyramid%20of%20Pain%20_%20Sqrrl.pdf.

[56] S. Commerce, "A "Kill Chain" analysis of the 2013 target data breach," in *The Target Store Data Breaches: Examination and Insight*, Jan. 2014, pp. 41–60.

[57] Davidjbianco, *Enterprise Detection & Response: The Pyramid of Pain*, Mar. 2013. [Online]. Available: https://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html.

[58] P. Ewing and D. Kerr, *The Endgame Guide to Threat Hunting*, en. [Online]. Available: https://cyber-edge.com/resources/the-endgame-guide-to-threat-hunting/.

[59] K. Scarfone, *The Hunters Handbook*.

[60] *GUIDE TO CYBER THREAT HUNTING*. [Online]. Available: https://www.tylertech.com/services/ndiscovery/nDiscovery-Threat-Hunting.pdf.

[61] "X-Force Threat Intelligence Index 2022," IBM Security, Tech. Rep., Feb. 2022. [Online]. Available: https://www.ibm.com/downloads/cas/ADLMYLAZ.

[62] S. Trent, R. R. Hoffman, D. Merritt, and S. Smith, "Modelling the Cognitive Work of Cyber Protection Teams," *The Cyber Defense Review*, vol. 4, no. 1, pp. 125–136, 2019, Publisher: Army Cyber Institute, ISSN: 2474-2120. [Online]. Available: http://www.jstor.org/stable/26623071.

[63] J. M. Spring and P. Illari, "Review of Human Decision-making during Computer Security Incident Analysis," en, *Digital Threats: Research and Practice*, vol. 2, no. 2, pp. 1–47, Jun. 2021, ISSN: 2692-1626, 2576-5337. DOI: 10.1145/3427787. [Online]. Available: https://dl.acm.org/doi/10.1145/3427787.

[64] Jason Chaffetz, Mark Meadows, and Will Hurd, "The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation," U.S. House of Representatives Committee on Oversight and Government Reform, Majority Staff Report, p. 241. [Online]. Available: https://oversight.house.gov/wp-content/uploads/2016/09/The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf.

[65] *Lessons to Learn from the OPM Breach*, en, Jun. 2015. [Online]. Available: https://www.tenable.com/blog/lessons-to-learn-from-the-opm-breach.

[66] R. Lemos, *5 Lessons Learned From OPM Data Breach*, en-US, Sep. 2016. [Online]. Available: https://www.eweek.com/security/5-revelations-from-opm-data-breach-report/.

[67] *Lessons from the OPM breach*, en, Sep. 2016. [Online]. Available: https://gcn.com/cybersecurity/2016/09/lessons-from-the-opm-breach/316728/.

[68] S. M. Kerner, *Lessons Learned from the OPM Breach eSecurity Planet*, en-US, Dec. 2017. [Online]. Available: https://www.esecurityplanet.com/threats/lessons-learned-from-the-opm-breach/.

[69] D. Bianco, *A Simple Hunting Maturity Model*, en, Blog. [Online]. Available: http://detect-respond.blogspot.com/2015/10/a-simple-hunting-maturity-model.html.

[70] R. Gutzwiller, J. Dykstra, and B. Payne, "Gaps and Opportunities in Situational Awareness for Cybersecurity," en, *Digital Threats: Research and Practice*, vol. 1, no. 3, pp. 1–6, Sep. 2020, ISSN: 2692-1626, 2576-5337. DOI: 10.1145/3384471. [Online]. Available: https://dl.acm.org/doi/10.1145/3384471.

[71] R. M. Lee and R. T. Lee, "SANS 2018 Threat Hunting Survey Results," en, SANS Institute, Tech. Rep., 2018.

[72] S. M. Milajerdi, B. Eshete, R. 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*, pp. 1795–1812, Nov. 2019, arXiv: 1910.00056. DOI: 10.1145/3319535.3363217. [Online]. Available: http://arxiv.org/abs/1910.00056.

[73] R. Wei, L. Cai, A. Yu, and D. Meng, *DeepHunter: A Graph Neural Network Based Approach for Robust Cyber Threat Hunting*, arXiv:2104.09806 [cs], Apr. 2021. [Online]. Available: http://arxiv.org/abs/2104.09806.

[74] P. Karuna, E. Hemberg, U.-M. O'Reilly, and N. Rutar, "Automating Cyber Threat Hunting Using NLP, Automated Query Generation, and Genetic Perturbation," *arXiv: 2104.11576 [cs]*, Apr. 2021, arXiv: 2104.11576. [Online]. Available: http://arxiv.org/abs/2104.11576.

[75] A. B. Ajmal, M. Alam, A. A. Khaliq, S. Khan, Z. Qadir, and M. A. P. Mahmud, "Last Line of Defense: Reliability Through Inducing Cyber Threat Hunting With Deception in SCADA Networks," *IEEE Access*, vol. 9, pp. 126 789–126 800, 2021, Conference Name: IEEE Access, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3111420.

[76] *Cyber Threat Hunting Solutions IBM*, en-us. [Online]. Available: https://www.ibm.com/qradar/threat-hunting.

[77] *Threat Hunting Tools*, en-US. [Online]. Available: https://www.cyrebro.io/threat-hunting/.

[78] A. J. Horta Neto and A. 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)*, Dec. 2020, pp. 1823–1830. DOI: 10.1109/BigData50022.2020.9378213.

[79] S.-X. Lin, Z.-J. Li, T.-Y. Chen, and D.-J. Wu, "Attack Tactic Labeling for Cyber Threat Hunting," en, p. 7,

[80] R. J. Chenail, "Interviewing the Investigator: Strategies for Addressing Instrumentation and Researcher Bias Concerns in Qualitative Research," en, p. 8, 2011.

[81] J. Saldaña, "The coding manual for qualitative researchers," *The coding manual for qualitative researchers*, pp. 1–440, 2021, Publisher: SAGE publications Ltd.

[82] G. Guest, K. MacQueen, and E. Namey, *Applied Thematic Analysis*, en. 2455 Teller Road,ăThousand OaksăCaliforniaă91320ăUnited States: SAGE Publications, Inc., 2012, ISBN: 978-1-4129-7167-6 978-1-4833-8443-6. DOI: 10.4135/9781483384436. [Online]. Available: https://methods.sagepub.com/book/applied-thematic-analysis.

[83] G. Guest, A. Bunce, and L. Johnson, "How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability," en, *Field Methods*, vol. 18, no. 1, pp. 59–82, Feb. 2006, ISSN: 1525-822X, 1552-3969. DOI: 10.1177/1525822X05279903. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1525822X05279903.

[84] J. Huck and F. Breitinger, "Wake Up Digital Forensics Community and Help Combat Ransomware," *IEEE Security & Privacy*, vol. 20, no. 4, pp. 61–70, Jul. 2022, Conference Name: IEEE Security & Privacy, ISSN: 1558-4046. DOI: 10.1109/MSEC.2021.3137018.

[85] J. Pettigrew and J. Ryan, "Making Successful Security Decisions: A Qualitative Evaluation," *IEEE Security & Privacy*, vol. 10, no. 1, pp. 60–68, Jan. 2012, Conference Name: IEEE Security & Privacy, ISSN: 1558-4046. DOI: 10.1109/MSP.2011.128.

[86] M. De Gramatica, F. Massacci, W. Shim, A. Tedeschi, and J. Williams, "IT Interdependence and the Economic Fairness of Cybersecurity Regulations for Civil Aviation," *IEEE Security & Privacy*, vol. 13, no. 5, pp. 52–61, Sep. 2015, Conference Name: IEEE Security & Privacy, ISSN: 1558-4046. DOI: 10.1109/MSP.2015.98.

[87] A. Ebrahimi, A. Leithner, E. A. Lowham, S. Tiscareño, M. Battle, M. Elmjouie, M. Vieira, and J. Watkins, "National Guard Cyber Protection Teams as a Response to Cybersecurity Threats," en, p. 70,

[88] *Cyber Flag 22: A multinational exercise aims to make defense networks safer and more secur*, en-US. [Online]. Available: https://www.mycg.uscg.mil/News/Article/3156685/cyber-flag-22-a-multinational-exercise-aims-to-make-defense-networks-safer-and/https%3A%2F%2Fwww.mycg.uscg.mil%2FNews%2FArticle%2F3156685%2Fcyber-flag-22-a-multinational-exercise-aims-to-make-defense-networks-safer-and%2F.

[89] *Senate confirms Jen Easterly as head of U.S. cyber agency*, en. [Online]. Available: https://www.politico.com/news/2021/07/12/senate-confirms-jen-easterly-cyber-499335.

[90] F. Araujo, D. Kirat, X. Shu, T. Taylor, and J. Jang, "Evidential Cyber Threat Hunting," *arXiv:2104.10319 [cs]*, Apr. 2021, arXiv: 2104.10319. [Online]. Available: http://arxiv.org/abs/2104.10319.

[91] *GIAC Find a Certification*. [Online]. Available: https://www.giac.org/certifications/.

[92] *Cybersecurity Certifications - Information Security Certifications - (ISC)š*. [Online]. Available: https://www.isc2.org:443/Certifications.

[93] C. Maslach and M. Leiter, "Burnout," en, in *Stress: Concepts, Cognition, Emotion, and Behavior*, Elsevier, 2016, pp. 351–357, ISBN: 978-0-12-800951-2. DOI: 10.1016/B978-0-12-800951-2.00044-3. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780128009512000443.

[94] EC-Council, *What Is the Pyramid of Pain, and Why Is It Important in Threat Detection?* en-US, Oct. 2022. [Online]. Available: https://www.eccouncil.org/cybersecurity-exchange/threat-intelligence/pyramid-pain-threat-detection/.

[95] "CISA Strategic Plan 2023-2025," en, p. 37,

[96] *MITRE ATT&CKő*. [Online]. Available: https://attack.mitre.org/.

[97] *The Cambridge Handbook of the Learning Sciences*, 2nd ed., ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2014. [Online]. Available: https://doi.org/10.1017/CBO9781139519526.

[98] M. Borg, "The apprenticeship of observation," *Elt Journal*, vol. 58, pp. 274–276, Jul. 2004. DOI: 10.1093/elt/58.3.274.

[99] P. Ryan, "Is apprenticeship better? a review of the economic evidence," en, *Journal of Vocational Education & Training*, vol. 50, no. 2, pp. 289–325, Jun. 1998, ISSN: 1363-6820, 1747-5090. DOI: 10.1080/13636829800200050. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/13636829800200050.

[100] S. Smith, E. Taylor-Smith, K. Fabian, M. Barr, T. Berg, D. Cutting, J. Paterson, T. Young, and M. Zarb, "Computing degree apprenticeships: An opportunity to address gender imbalance in the IT sector?" In *2020 IEEE Frontiers in Education Conference (FIE)*, ISSN: 2377-634X, Oct. 2020, pp. 1–8. DOI: 10.1109/FIE44824.2020.9274144.

[101] G. Stoker, U. Clark, M. Vanajakumari, and W. Wetherill, "Building a Cybersecurity Apprenticeship Program: Early-Stage Success and Some Lessons Learned," en, p. 10, 2021.

[102] L. Williams and R. R. Kessler, *Pair Programming Illuminated*, en. Addison-Wesley Professional, 2003, Google-Books-ID: LRQhdlrKNE8C, ISBN: 978-0-201-74576-4.

[103] V. R. Basili, "Models and metrics for software management and engineering," Tech. Rep. NASA-CR-182953, Jan. 1988, NTRS Author Affiliations: Maryland Univ. NTRS Document ID: 19880014816 NTRS Research Center: Legacy CDMS (CDMS). [Online]. Available: https://ntrs.nasa.gov/citations/19880014816.

[104] A. Zibak, C. Sauerwein, and A. C. Simpson, "Threat Intelligence Quality Dimensions for Research and Practice," en, *Digital Threats: Research and Practice*, vol. 3, no. 4, pp. 1–22, Dec. 2022, ISSN: 2692-1626, 2576-5337. DOI: 10.1145/3484202. [Online]. Available: https://dl.acm.org/doi/10.1145/3484202.

[105] V. G. Li, M. Dunn, P. Pearce, D. McCoy, G. M. Voelker, S. Savage, and K. Levchenko, "Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence," en,

[106] S. Bromander, M. Swimmer, L. P. Muller, A. Jøsang, M. Eian, G. Skjøtskift, and F. Borg, "Investigating Sharing of Cyber Threat Intelligence and Proposing A New Data Model for Enabling Automation in Knowledge Representation and Exchange," en, *Digital Threats: Research and Practice*, vol. 3, no. 1, pp. 1–22, Mar. 2022, ISSN: 2692-1626, 2576-5337. DOI: 10.1145/3458027. [Online]. Available: https://dl.acm.org/doi/10.1145/3458027.

[107] L. F. DeKoven, A. Randall, A. Mirian, G. Akiwate, A. Blume, L. K. Saul, A. Schulman, G. M. Voelker, and S. Savage, "Measuring Security Practices and How They Impact Security," in *Proceedings of the Internet Measurement Conference*, ser. IMC '19, New York, NY, USA: Association for Computing Machinery, Oct. 2019, pp. 36–49, ISBN: 978-1-4503-6948-0. DOI: 10.1145/3355369.3355571. [Online]. Available: https://doi.org/10.1145/3355369.3355571.

A. INTERVIEW PROTOCOL

This interview guide is organized into headings that identify the main themes or lines of inquiry, under which individual questions to be asked are given. The main questions are listed first, with possible follow-on probes nested underneath. The follow on questions were optional and sometimes not used if they were not applicable or already addressed by the subject. Due to the nature of a semi-structured interview, the questions may not be asked exactly as written or in the same sequence, but may be adjusted depending on how the conversation flows. The intention is that they serve mainly as a reminder or checklist for the interviewer.

*: indicates that the question was added after a theme emerged from the first few interviews.

A.1 Introduction (~ 10 mins)

After 2 administrative reminders, this portion of the interview will be focused on discerning the history of the interviewee and enumerating the teams on which they were a member. If they were threat hunters for multiple organizations then the other sets of questions may be able to be asked about multiple organizations with the understanding that some of the information may be outdated.

A.1.1 Reminders:

- This interview is being conducted over an insecure medium. Please keep all discussion limited to only Unclassified information and refrain from providing customer details that should remain confidential.
- This interview is being recorded.
- Do you have any questions about this interview's classification policy or any general questions about this interview, the study, the data being collected or the researchers?

A.1.2 Questions:

- How long have you been on your current threat hunt team?
  - How many missions have you been engaged on?
  - How many of your missions successfully identified adversary activity?
- Have you worked on any threat hunt teams previously?
  - Which ones? For how long?
  - How many missions were you deployed on with your previous team?
  - How many of those missions successfully identified adversary activity?

A.2 Threat Hunting Processes (~ 30 mins)

This portion of the interview focuses on the process used by an agency and how that process came about, how effective it is, how it could be improved, etc. This will primarily contain questions relevant to research question 1.

A.2.1 Questions:

- Could you please draw a diagram of the current threat hunting process that is followed by your team and explain the diagram?
  - Is this process standardized across all teams in your agency? If no, why not?
  - How often is the process modified?
  - What causes the process to be modified?
  - How strictly is the process followed?
  - Do you believe this is too strict or to lose? Why?
  - What do you find to be the most problematic parts of this process? Why?
- What models are incorporated in your process? An example model would be the OODA loop.

- In what way do you incorporate the following models into your process:
  - Kill chain Model?
  - Pyramid of Pain?
  - MITRE ATT&CK?
  - Hypothesis creating and checking?
    * Can you give us an example of a hypothesis?
    * How are they documented and shared with the team?
- How is the process tracked or process documentation accessed when on engagement?
  - How often is it referenced?
  - How granular is the process documentation?
  - Why that level of granularity?
  - Do you believe it is too specific or vague? Why?
- How was the current process developed?
  - Were any alternative processes or frameworks considered?
  - Who was given the opportunity to provide input in the development and institution of the process?
- Can you think of a mission where the TH process was not as helpful as you would have liked or even counter productive? (Question should be sent ahead of time so the subject have time to remember an event)
  - Please describe the event to without giving us details about the customer.
  - How long did it take before the process deficiencies were made obvious?
  - What was the shortcoming?
  - What changes to the process would help alleviate this (these) shortcoming(s)?
  - How common are experiences like this?

  - Did the experience level of the team contribute to the issue?

- Can you think of a second mission where the TH process was not as helpful as you would have liked or even counter productive? (Question should be sent ahead of time so the subject have time to remember an event)
  - Please describe the event to without giving us details about the customer.
  - How long did it take before the process deficiencies were made obvious?
  - What was the shortcoming?
  - What changes to the process would help alleviate this (these) shortcoming(s)?
  - How common are experiences like this?
  - Did the experience level of the team contribute to the issue?

- How much of the process is automated?
  - Who automated it?
  - How much more of it can be automated?
  - Why hasnt it been?

- When adversary activity is typically found on a Threat Hunt, how is it typically found?
  - How experienced is the member that typically finds it?

- Roughly how often is adversary activity found:
  - With a sensor alert?
  - With a tip from another team conducting an Incident Response mission?
  - With a tip from intelligence sources?
  - Another way?
    * When the activity is discovered another way, is there a common way or location it is discovered?

        * How often is the person who discovers the activity another way an inexperienced member?

- Does your process include a mission plan?*
  - Is it helpful?*
  - Does it ever get in the way of operations?*
  - What would make it more helpful?*

A.3 Integration of New Members ($\sim$ 20 mins)

This portion of the interview focuses on how newer members are affected by a change in process and what improvements could be made to onboard new members more efficiently. These questions correspond primarily to research questions 2 and 3.

A.3.1 Questions:

- How long does it take new members of the team to independently reason through the current process without help?
  - Why does it take this long?
  - Are there variables that speed this up or slow it down?
  - Does this change if the new member has threat hunt experience?
  - Does this change if the new member has other experiences in their background? What experiences are helpful?
  - How long does it take new members to become a net positive for the team when on engagement?
  - Does this happen at the point they can reason through the current process independently or at a different time?[1]
    * Why do you think this is?

[1]↑This question and follow-up question were removed after 5 interviews

- What is a good measure of a member's expertise?*
  - Experience?*
  - Certifications?*
  - Personality?*
  - Willingness to put in own time?*
  - Is this different for potential?*
- Have you observed any changes to the process that allow for a faster integration of new members?
  - What were they?
  - Were these changes made permanent?
  - If no, why not?
  - To what extent have other team leads observed a similar phenomenon?
  - Would you have any other recommendations?*
- Have you observed any changes that slow down the integration of new members?
  - What were they?
  - Were these changes made permanent?
  - To what extent have other team leads observed a similar phenomenon?
- Before we close, are there any additional questions that I should have asked but didn't?
- Is there anything else I should know about threat hunting processes?

A.4 Closing (2 mins)

This is simply an opportunity to thank the participant and bring the interview to a close. No closing questions will be asked.

B. OTHER RESULTS

The two result sections in this appendix were not included in RQ1.2 because fewer than 4 subjects mentioned these issues. Both of these sections were individually interesting and therefore have been placed into this appendix.

B.1 Timing of process components

B.1.1 Problems

In addition to the process timing being important for sensor placement, subjects described other issues with timing. The process governs when the deployment and recall of teams occurs. At some organizations, it also determines how long a team will spend baselining when they first arrive. Some teams used the same engagement length on all engagements and did not modify the length of time based on deployment. This created scenarios where subjects were given multiple weeks to hunt on a very small network or teams were required to hunt a very large network in an unreasonably short amount of time. Three example issues that were provided by three separate subjects ranging from analysts to team leads indicated that they experienced missions where the process timing led to inefficiencies. One of these three subjects said "*So the process we had, I said, we go on a mission for two weeks, right? ... if your network is super small and you're still there for two weeks, you kinda don't really have much else to do. ... I feel like we could have done a little more like investigating after we were done with the spreadsheet, [but] nobody cared.*" The second concurred, saying: "*we were sitting around fiddling our thumbs by the end of the second week because it was such a small environment*". The last subject thought the baselining phase of the mission lasted too long: "*it wasn't like we need to spend two to three days dedicated to doing this [baselining]. It could have been like a half day and we're good.*"

B.1.2 Solutions

Of the three subjects that provided example issues where the process timing was too rigid and all three subjects provided suggested solutions. One of these subjects was an analyst

who suggested: "*I would say deploy sensors before the team gets there. Have it collected the data and then, Hmm. I feel like you wouldn't even need to be on site. You could just remote into your system, do it. Okay? And I feel like it should be drawn out and not just one single time, you know?*" Another of these three was a team lead who said their team was already working toward such a solution: "*Right now it's just like a strict two weeks onsite and depending on the size of the network, that may be more than enough, which has happened and sometimes [its] not enough at all. Right. And that's just solely based on the size of the network and how much data that we're looking at. That's the other [thing] that we're working on ... [in the future] we [will] have the whole month to work with the partner ... So if we did need three weeks on-site or the whole month on site, oh, I hope that would never happen — that would get kind of exhausting ... but yeah, if we needed it, we could do that. So that's in the works.*" This change had already occurred at the other two organizations. None of the other subjects from any organization specifically discussed flexibility in timing being helpful, but this one organization was the only organization that experienced such timing issues.

The third subject was a team lead and believed the baselining step was too long. They discussed that they were able to leverage tools to get a good baseline in a few hours but due to the process, they still had to spend the full 2-3 days baselining. "*I think it takes too long and what I was doing during the last hunt was trying to ... [explains using a tool to baseline] ... and having that be automatic and immediate, then you can kind of shift the window baseline out from two to three days where during that period before it was basically, let's just look around at network traffic and write in [notes] what it looks like or try and reference their network diagram they gave us. Now it moves to a five-minute thing where it's all set up in [tool].*" The automation aspect of this example will be further discussed in §4.2.1 but this example is relevant here because the team lead wanted to be able to adjust the time spent baselining and due to the process was not able to. Not every team will be capable of achieving a good baseline quickly, either due to the complexities of a customer network or due to the expertise of the team. In these situations, if the team lead can adjust how long phases like baselining take, it might help make the process less of a hindrance.

B.2 Cloud

B.2.1 Problems

In interviews, two subjects mentioned hunting on cloud platforms. One subject mentioned having dedicated personnel for special situations like that. While discussing expert analysts they said: "*We may be asked to do cloud on-prem or ICS so we have to have people in the toolbox to kind of pull and put on these [missions] that are either willing to learn or know something about ... those.*" A second subject seemed less confident in dealing with the cloud. They were asked about situations where their mission plan would not be helpful. They said: "*it's not helpful in that sometimes ... if there are unique situations, and one of these would be like in the cloud based area, because so many are moving towards using the cloud at least in part of their environment, that we don't really know what to do with that and so that's like definitely a deficit in the mission plan. We just don't touch it basically.*"

Interviewer: "*So the mission plan just says don't touch it?*"

Subject: "*I mean it's on there but the other thing is that it usually always involves a third party, whoever you're using for your cloud environment and that's why we've never really actually had to deal with it. Is, you know, working through any kind of third party agreement — usually they don't want you to so that's a legal issue.*" As the subject mentions, many organizations are moving operations to the cloud. Not being able to deal with situations where a third party is involved seems to be a deficiency TH capability.