NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

INSIDER THREAT DATA SHARING

by

Jeremey J. Sellen

September 2016

Thesis Advisor: Shelley Gallup
Co-Advisor: Thomas Anderson

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE September 2016	3. REPORT TYPE AND DATES COVERED Master's thesis

4. TITLE AND SUBTITLE INSIDER THREAT DATA SHARING	5. FUNDING NUMBERS
6. AUTHOR(S) Jeremey J. Sellen	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

Insider threat is a significant problem for both governmental and private organizations. Individuals can do immense harm with their trusted accesses. To combat this threat, organizations have created departments with trained analysts whose sole purpose is to deter, detect, and mitigate the insider threat. These analysts monitor employees and analyze activities to detect dangerous practices, whether witting or unwitting, and report these actions to supervisors for mitigation.

When organizations share insider threat information among each other, it can improve all organizations' abilities to deter, detect, or mitigate the insider threat. The challenge lies in merging external and existing data with as little human interaction as possible. This thesis examines the work that takes place in an insider threat department and identifies requirements for a solution that would allow for information sharing between organizations.

14. SUBJECT TERMS insider threat, data sharing, hub operations, heterogeneous data	15. NUMBER OF PAGES 93
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

INSIDER THREAT DATA SHARING

Jeremey J. Sellen
Lieutenant, United States Navy
B.S., Old Dominion University, 2009

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN CYBER SYSTEMS AND OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
September 2016

Approved by: Shelley Gallup, Ph.D.
Thesis Advisor

Thomas Anderson, Ph.D.
Co-Advisor

Cynthia Irvine, Ph.D.
Chair, Cyber Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Insider threat is a significant problem for both governmental and private organizations. Individuals can do immense harm with their trusted accesses. To combat this threat, organizations have created departments with trained analysts whose sole purpose is to deter, detect, and mitigate the insider threat. These analysts monitor employees and analyze activities to detect dangerous practices, whether witting or unwitting, and report these actions to supervisors for mitigation.

When organizations share insider threat information among each other, it can improve all organizations' abilities to deter, detect, or mitigate the insider threat. The challenge lies in merging external and existing data with as little human interaction as possible. This thesis examines the work that takes place in an insider threat department and identifies requirements for a solution that would allow for information sharing between organizations.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
- A. BACKGROUND 1
- B. PROBLEM STATEMENT 3
- C. RESEARCH QUESTIONS 3
- D. SIGNIFICANCE OF RESEARCH 4
- E. STRUCTURE OF THIS THESIS 4

II. LITERATURE REVIEW 5
- A. INSIDER THREAT 5
- B. INSIDER THREAT ANALYSIS 9
- C. DATABASES 16

III. METHODOLOGY 25
- A. THE MODEL 25
- B. STRUCTURED SYSTEMS ANALYSIS 28
- C. METHOD 30

IV. ANALYSIS 33
- A. ENVIRONMENT 34
  - 1. Shore-Based (Domestic) 34
  - 2. Shore-Based (Foreign) 35
  - 3. Afloat Platforms 35
  - 4. Undersea 36
  - 5. Airborne 37
  - 6. Forward-Deployed 37
- B. PEOPLE 38
  - 1. IT Support Personnel 38
  - 2. INT Analyst 40
  - 3. INT Supervisor 41
  - 4. Subject Matter Expert (SME) 42
  - 5. Senior Official 43
- C. TECHNOLOGY 43
  - 1. Network Nodes 44
  - 2. UAM Clients 44
  - 3. UAM Servers 47
  - 4. Case Management 47
- D. PROCESSES 48

E. WORKFLOW 50
1. Recording / Automated Analysis Phase 51
2. Manual Analysis Phase 52
3. Case Management Phase 55
4. Administrative / Legal Action Phase 57
F. DFD 58
G. ARTIFACTS OF WORK 61

V. CONCLUSION AND FUTURE WORK 63

APPENDIX. DSS RESPONSE TO INTERVIEW QUESTIONS 65
A. SYSTEMS 65
B. DATA 66
C. AUTOMATION AVENUES 66
D. INFORMATION SHARING AND PROCESS 67
E. ANALYST / WORKFLOW 67
F. CUSTOMER NEEDS 69
G. IT PERSONNEL QUESTIONS 70

SUPPLEMENTAL 71
A. INT HUB WORKFLOW 71
B. INT HUB DATA FLOW DIAGRAM 71

LIST OF REFERENCES 73

INITIAL DISTRIBUTION LIST 75

LIST OF FIGURES

Figure 1. Average Annualized Cost Based on Frequency from 252 Global Companies. Source: Ponemon Institute LLD (2015).....2

Figure 2. Possible Cyber Observables from Insider Actions. Source: Brackney & Anderson (2004). 7

Figure 3. INSA Insider Threat Road-Map. Adapted from Balakrishnan (2015). 10

Figure 4. Sliding Scale of Cybersecurity. Adapted from Lee (2015).....12

Figure 5. The Intelligence Process. Source: Lee (2015). 14

Figure 6. Sequence of Events in an Insider Threat Case. Source: Brackney & Anderson (2004). 15

Figure 7. Example Database Tables 17

Figure 8. Saving All Persons as One Entity Type. Source: Ambler (2000).....22

Figure 9. Using Primary-Foreign Keys for Multiple Sub-Classes of Person. Source: Ambler (2000). 22

Figure 10. Ecosystem Approach to Insider Threat Sharing Problem.....26

Figure 11. Example Cross-Functional Workflow Diagram 29

Figure 12. Example DFD 30

Figure 13. Triad of Work Produced within an InT Hub.....49

Figure 14. Initial Recording / Automated Analysis Phase of InT Hub Workflow.....51

Figure 15. Workflow of Manual Analysis.....54

Figure 16. Case Management Phase.....56

Figure 17. Administrative / Legal Action Phase 58

Figure 18. Level 0 DFD for InT Hub Operations.....59

Figure 19. Level 1 DFD of InT Hub Operations.....60

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1. SQL Query Result 18

Table 2. SQL Query Result 19

Table 3. Work and Artifacts of Work for IT Support Personnel 40

Table 4. Work and Artifacts of Work for InT Analyst 40

Table 5. Work and Artifacts of Work for InT Supervisor 42

Table 6. Work and Artifacts of Work for SME 43

Table 7. Work and Artifacts of Work for Senior Official 43

Table 8. Work and Artifacts of Work for UAM Clients 45

Table 9. Work and Artifacts of Work for UAM Servers 47

Table 10. Work and Artifacts of Work for Case Management Tool 48

Table 11. Work and Artifacts of Work for Processes 50

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACID	Atomicity, Consistency, Isolation, Durability
CI	Counter Intelligence
CI&A	Confidentiality, Integrity, and Availability
DFD	Data Flow Diagram
DITMAC	DOD Insider Threat Management Analysis Center
DOD	Department of Defense
DON	Department of the Navy
DSS	Defense Security Service
FBI	Federal Bureau of Investigation
HR	Human Resources
IDPS	Intrusion Detection/Prevention System
ISP	Internet Service Provider
IT	Information Technology
MMOWGLI	Massive Multiplayer Online Wargame Leveraging the Internet
NGA	National Geospatial-Intelligence Agency
NPS	Naval Postgraduate School
NSA	National Security Agency
PFC	Private First Class
PII	Personally Identifiable Information
RAM	Random Access Memory
RDBMS	Relational Database Management System
SE	Social Engineering
SIEM	Security Information and Event Management
SME	Subject Matter Expert
SQL	Structured Query Language
TTP	Tactics, Techniques, and Procedures
UAM	User Activity Monitoring
USG	U.S. Government
USN	United States Navy
VSAT	Very Small Aperture Terminal

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Nothing of great importance can be accomplished without the support of family. My wife and children were very patient with my long nights and weekends generating this thesis. They, above all others, deserve the utmost gratitude for providing me the moral support needed to complete this work.

A huge thank-you goes out, as well, to my advisor and co-advisor, Dr. Shelley Gallup and Dr. Thomas Anderson, respectively, for their guidance, expertise, and patience in this long process. Without them, I would have had a very rough road indeed.

To my other NPS insider threat team members, it was an honor to work with all of you. Ryan Kelly, Rich Mascolo, Scott McKenzie, Will Campbell, and Dr. Nelson Irvine: you all made it an enjoyable ride.

The members of my cohort and their amazing personalities offset the stress of the demanding academic experience. Cayanne, Dan, Ian, John, Lindsey, Lou, Max, Mel, Melissa, Roman, Steve, and Welton, thank you all for always keeping the days light-hearted and fun.

A special thank-you goes out to those members of the DITMAC, DSS, NGA, Coast Guard, NITTF, Navy, and Army who took time out of their busy days to meet with me. They provided valuable data that I needed to gain a better understanding of the nature of the insider threat program within the DOD.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Insider threats (InT) pose a significant risk to organizations. Because most organizations' security practices focus on external malicious threats, insiders—with their trusted access within an organization—can cause tremendous damage. . The insider's malicious acts are often mistaken as normal operations, making it difficult to discern bad actions from ordinary daily activity. Compounding the problem, accidental or unintentional acts by unwitting employees or partners can also have a significant impact on an organization's security.

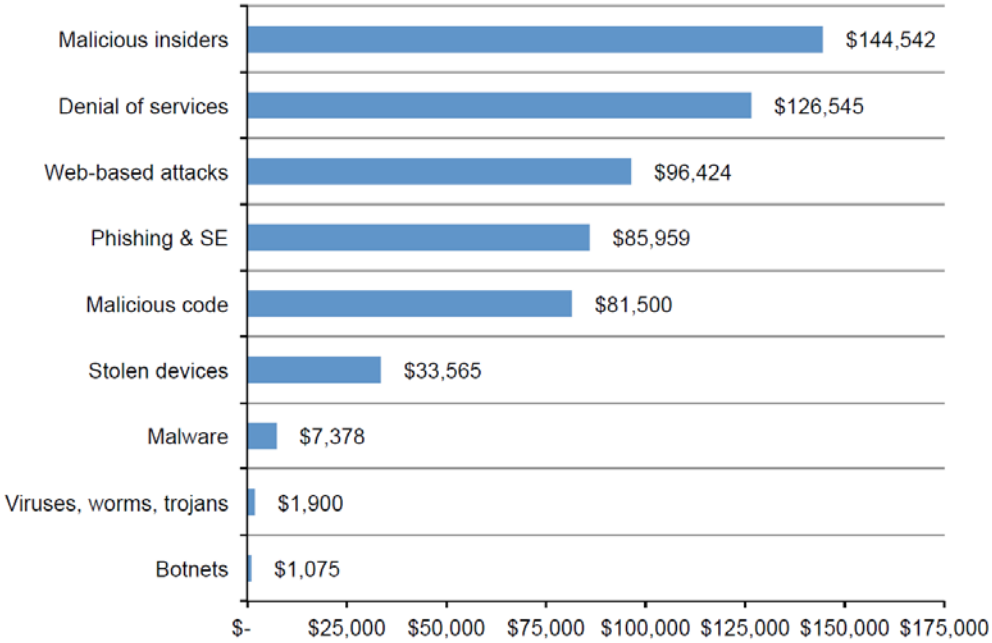
Since these threats are so diverse in ability and intent, an organization's InT analysts operate at a disadvantage. They must analyze what appears to be everyday business activity to discern the actions that stand out as malicious or otherwise dangerous. A shared understanding between organizations of the actual threats and methods used to deter, detect, and mitigate the InT could have a significant positive effect. This thesis explores the nuances of the insider threat, the InT analysis process, and technological challenges, all with the focus of identifying the requirements of a solution for sharing key InT detection and case information between InT programs. It is worth noting the word "case" in this context is not referring to an investigative case, but an inquiry case within an InT hub.

A. BACKGROUND

Insider threats are not limited to government or non-government organizations, nor are they limited to a specific region of the world. The insider threat can exist in any organization, anywhere, at any time. Negative effects the InT can have on an organization's security range from data corruption/loss to violence in the workplace. The ability of the insider to have a negative impact on an organization's people or information while exploiting their trusted access can be immense.

There is documentation that captures the gravity of this threat. The Ponemon Institute researched 252 companies from countries across the globe and found insider threats caused the highest average annualized cost of cybercrime to organizations

(Ponemon Institute LLD, 2015). Figure 1 shows the ranking of the top causes of cybercrime for 2015.



Malicious insiders are responsible for the highest average annualized cost based on the frequency of attacks related to cybercrime occurring at companies around the globe (SE stands for social engineering).

Figure 1. Average Annualized Cost Based on Frequency from 252 Global Companies. Source: Ponemon Institute LLD (2015).

Due to the severe impact of insider threats, the U.S. government has made significant strides to address this issue. In November of 2012, President Obama released a Presidential Memorandum directing all executive level departments to establish the ability to collect, analyze, fuse, and respond to InT-related matters (Obama, 2012). As a response to this directive, the Department of Defense (DOD) instructed all services to adopt an InT program (Department of Defense [DOD], 2014). The chosen architecture for this InT program is a hub structure with all data coming into one centralized department for analysis and processing (R. Laylo, interview with author, March 14, 2015). This allows for tight control of access and authority of the InT analysts as well as protecting hub operations data from the monitored employees.

Organizations implementing an InT program collect and store all employees' cyber activity information in some type of file structure. The simplest means for storing this information is in a flat file, like a log file. A more advanced solution is to store the cyber activity data within a User Activity Monitoring (UAM) solution. All of the UAM software solutions researched use a relational database for its data storage. A key point to consider is that not every organization will use the same data structure to store insider threat indicator and case data information. This diversity creates a technical challenge in sharing information between insider threat hubs.

B. PROBLEM STATEMENT

The U.S. Navy is in the initial phases of building an insider threat "hub" operations center capability. InT operations can benefit from sharing data between multiple InT organizations and between divisions and people. Currently, the DOD InT solution is a diverse set of technologies, methodologies, and policies that can be barriers to integration with other organizations within the DOD. The U.S. Navy is likely to experience these same barriers to integration. This research examines the insider threat from a technical perspective, analyzes InT hub workflows and relational database technology, and examines technologies available to assist in sharing data. In doing so, this thesis proposes a solution that will allow for improved InT hub operations.

C. RESEARCH QUESTIONS

The research hypothesis is that by sharing event and case data among hubs using potential technology-based solutions that enable knowledge flow, InT hubs can become more effective in deterring, detecting, and mitigating insider threats. The three research questions are:

1. What is the known set of cyber methods used by insiders?
2. What workflows are necessary to InT hub operations that require integrated data within and between hubs?
3. What technical means are required in order to share information between hubs?

D. SIGNIFICANCE OF RESEARCH

In June of 2015, the Naval Postgraduate School (NPS) Insider Threat Working Group conducted a Massive Multiplayer Online Wargame Leveraging the Internet (MMOWGLI) event. This was a crowdsourcing experiment to discover the best way to establish an insider threat hub for the U.S. Navy. This experiment, revealed that there was a need for information sharing among InT hubs to improve the ability of analysts to deter, detect, and mitigate the insider threat (Mascolo, 2016).

Upon the completion of this research, organizations, internal or external to the DOD, can consume the research findings to be able to create a collaborative synergy for their insider threat programs. The methods provided will attempt to encapsulate a solution tailored to InT operations. The Navy will benefit, as the product of this research will provide InT decision-makers both appropriate technologies and a road map to generate an effective InT mitigation architecture that allows for collaboration with other InT hubs.

E. STRUCTURE OF THIS THESIS

The structure of the remaining chapters is as follows:

- Chapter II provides a literature review in the area of insider threat, analyst workflow, and relational database technology.
- Chapter III describes the methodology and model used for providing a framework for this research, as well as a definition of a structured systems analysis and the method used for this research.
- Chapter IV provides an analysis of data gained from multiple InT organization interviews regarding workflows and data types to facilitate information sharing among InT hubs.
- Chapter V provides for a conclusion and recommendation for future work in the field of information sharing among InT hub organizations.

II. LITERATURE REVIEW

To provide a framework for the insider threat information-sharing problem, there are three key elements. The first element is an understanding of the cyber actions performed by insider threats that are observable via automated means. The second element is an understanding of the human and automated processes of InT hub monitoring, analysis, and reporting. The third element is an understanding of the relational database technology used to store InT information. This will permit better definition of how InT hubs can share this information.

A. INSIDER THREAT

The insider threat has two primary manifestations: witting and unwitting. Claycomb et al. define the witting, or malicious, insider as follows:

> A current or former employee, contractor, or other business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems. (Claycomb, Huth, Flynn, McIntire, & Lewellen, 2012)

The definition of "unwitting" is inadvertent or unintentional. The difference between these two types of actors is intent. The resulting state from their actions can be the same.

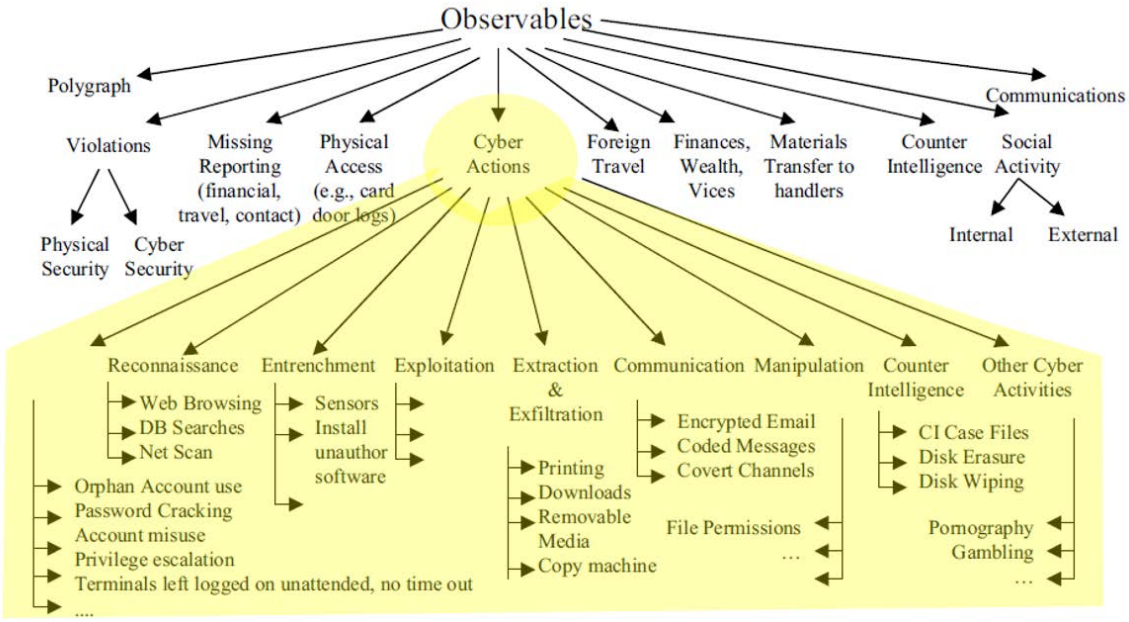
The InT is a significant threat to an organization's most sensitive information regardless of the actual intent. In 2015, Intel Security[1] conducted a survey of representatives from 1,155 worldwide organizations and discovered 43% of data loss cases were due to insider threats, where half were intentional and half were unintentional (Intel Security, 2015). Another 2015 study, the Clearswift Insider Threat Index, surveyed more than 500 IT supervisors and 4,000 employees from the U.S., U.K., Germany, and

[1] As a result of its acquisition of McAfee, Intel established Intel Security as an independent company on 7 September 2016 (Intel Security, 2016).

Australia. They reported 78% of breaches were due to insider influences, with 62% unwitting or accidental in nature (Masters, 2015)

These two studies illustrate two important themes. First, that insiders are responsible for a significant percentage of the data loss for organizations, and, second, that unwitting actions cause about half of those cases. This highlights both the pervasiveness of the InT as well as the need to minimize accidental cyber threats.

From a technology-driven perspective, it is important to examine how organizations can observe these InT cyber activities. Some of the current general-purpose methods to monitor insider cyber activity include Data Loss Prevention (DLP) programs, web-monitoring, keystroke or voice capturing, foreign travel tracking, daily time reporting, email monitoring, badge access monitoring, and log aggregation (Intelligence and National Security Alliance, 2013). In a 2004 conference workshop, the RAND Corporation performed a research study supporting multiple stakeholders (Brackney & Anderson, 2004). Specifically, the study examined elements of insider threats within the intelligence community and which actions were observable. The participants broke down the cyber observables resulting from an insider action, as shown in Figure 2.



This figure shows common cyber actions that are observable. This information can provide guidance for organizations to calibrate their monitoring tools to have the best ability to detect malicious activities.

Figure 2. Possible Cyber Observables from Insider Actions. Source: Brackney & Anderson (2004).

The list of actions in Figure 2 shows that different actions are observable within information technology (IT) systems and networks. The activities are all indicators that an organization can configure a UAM or other cyber monitoring solution to detect. Rules can filter for these indicators within the UAM and present them in an orderly manner to an InT analyst.

Knowledge of the observable acts is important, but so is knowledge of which methods insider threats are using. A 2012 Carnegie Mellon guide describes some common observables associated with insider threats, including:

- large files attached to emails,
- export of tagged documents (not meant to leave a network) outside a network,
- printing, copying, or downloading documents containing information such as PII or codenames,

- copying documents to removable media,
- remote accesses to company data, and
- sending emails to competitors, foreign governments, or personal email accounts. (Cappelli, Moore, & Trzeciak, 2012)

These examples show the threat uses diverse techniques, ranging from electronic to hard-copy data loss. Within the DOD, many cases of cyber espionage have resulted in data loss. Two preeminent cases were Edward Snowden and Private First Class (PFC) Bradley Manning. Upon examining both cases, this research defines these as instances of data loss due to poor application of monitoring solutions.

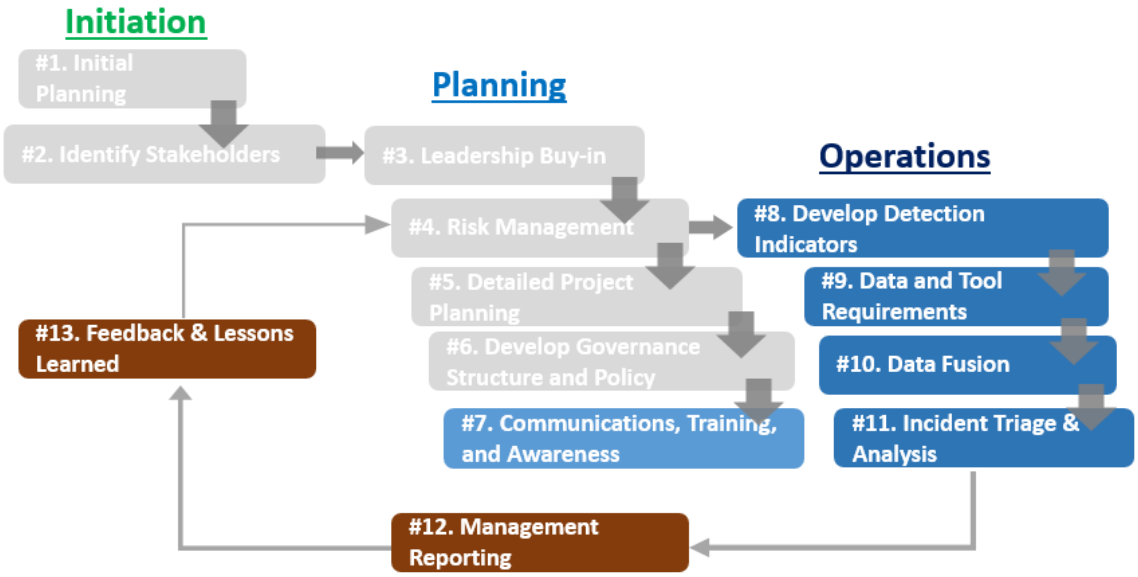
Edward Snowden was an infrastructure analyst for the National Security Agency (NSA) who worked as a Booz Allen Hamilton contractor to the NSA. He fled his station in Hawaii and flew to Hong Kong in May of 2013 with “a collection of highly classified documents” (Finn & Horwitz, 2013). According to an NBC news article, Snowden had system administrator privileges. With this access, he was reportedly able to use thumb drives to exfiltrate data with little to no oversight (Esposito & Cole, 2013). Organizations have the capability to configure UAM solutions to detect the use of thumb drives, but the NSA did not apply this capability correctly, as is evidenced by Snowden's ability to exfiltrate the data.

A second espionage case resulting in data loss was that of PFC Bradley Manning. Manning, an Army private, used his access to sensitive government computer networks to exfiltrate classified information and provide it to WikiLeaks. According to a *Washington Post* article from 2011, Manning was mentally unstable, had actively spoken out against the methods used by the U.S. Army, and was well-versed in computer technology and programming (Nakashima, 2011). This combination of factors is a classic recipe for a potential insider threat. Furthermore, a lack of proper security measures on the classified networks at his remote site in Iraq provided a means by which Manning could exfiltrate the data (Nakashima, 2011). This is yet another example of a failed application of a user monitoring capability.

Although Snowden and Manning provide high-impact, high-visibility insider threat cases, many other threats exist today. The means and methods used to bypass information system security in order to corrupt or remove information are almost endless. The threat vectors against networks and systems are plentiful because of advances in technology, regularly discovered software and hardware vulnerabilities, and zero-day attacks. This makes the job of InT analysts both difficult and important.

B. INSIDER THREAT ANALYSIS

To understand how organizations mitigate the insider threat problem, it was important to examine the process of insider threat analysis. InT analysis is a specialized process that implements many of the aspects of the intelligence process, described more in-depth later within this chapter. An understanding of the automated processes and human actions in this InT analysis overall system was important before attempting to improve upon that system. Figure 3 shows the roadmap created by the Intelligence and National Security Alliance to establish an InT program. Blocks #7 through #13 encompass the automated and human roles within the InT hub. The analyst and the process of fusing data to create intelligence regarding a possible InT are only a piece of the entire process. For the purposes of this research, the analysis and reporting pieces were of utmost importance as they contain the input and output processes required in the sharing of information between hubs.



The non-greyed out sections are the focus of the automated and human insider threat detection processes within the hub that can contribute to improved operations by sharing with other hubs. They are the part of the larger picture examined within this thesis.

Figure 3. INSA Insider Threat Road-Map. Adapted from Balakrishnan (2015).

Once the scope of the research excluded some of the administrative overhead of the establishment of a hub (greyed out in Figure 3), this left the framework for this section. The Communications, Training, and Awareness step (#7) is an important part of educating both employees and InT hub members on cybersecurity and common good practices (Balakrishnan, 2015). Knowledge from other InT hubs could be beneficial in creating these training aids to show both the good and bad behavior, which can assist in minimizing the unwitting insider threat.

The Develop Detection Indicators step (#8) allows stakeholders in the organization to identify and prioritize the information (Balakrishnan, 2015). They can then create indicators within the rulesets of monitoring tools to help detect insider threat activity. Without a thoughtful analysis of what is important, the data protection would become a matter of "protect everything" which would overload both the systems and the analysts.

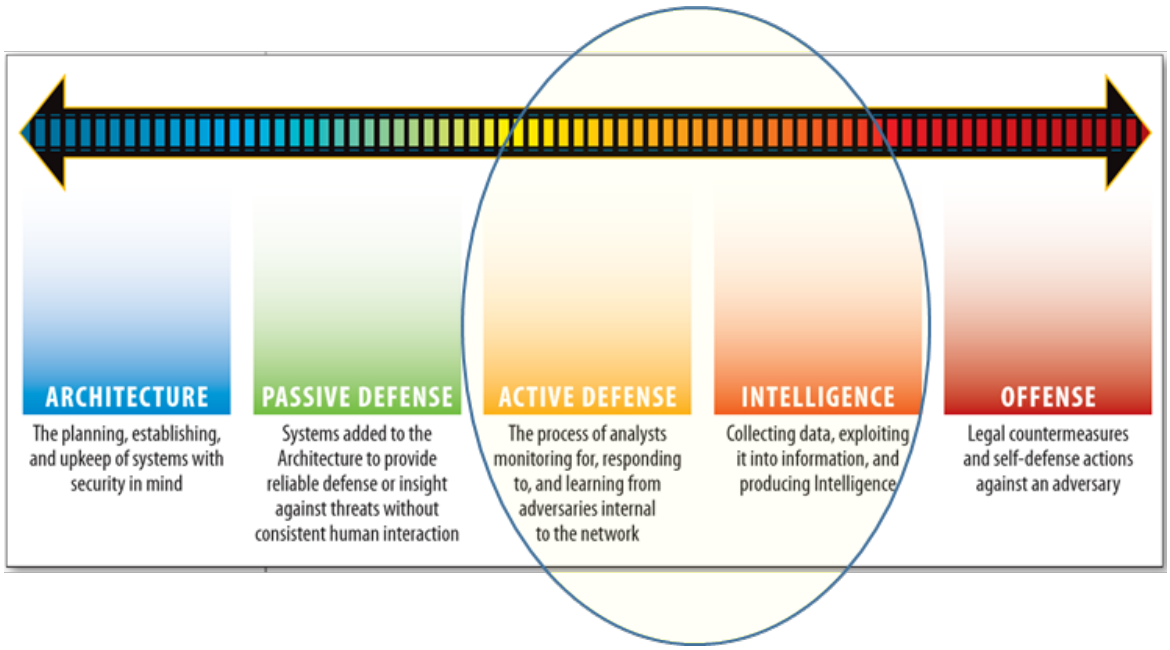
In the Data and Tool Requirements step (#9), an organization compares the requirements and indicators from the previous step with the available data and tool capabilities (Balakrishnan, 2015). If the right data or tool is not available, stakeholders must make a risk-based decision whether to allow a blind spot in user monitoring or to invest in research or further tool or information integration. The information and tool set support the Data Fusion step (#10). The data stores (i.e., databases, log files, etc.) must be combined with methods to monitor and report on the access and manipulation of information. A tool, such as a Security Incident Event Management (SIEM), may examine multiple events and elicit a relationship between the events within this data fusion step. This automated analysis can save an analyst some work in performing the next step.

The Incident Triage and Analysis step (#11) involves the analyst doing the analysis of the data. Analysts must examine the information presented by the UAM system to discern whether the activity is malicious. This involves looking at not only the indicator data, but also information from other data sources that can help put the activity into context. Balakrishnan stated a common example of a false negative is an alert from the UAM indicating large file exfiltration when a hard-working employee moves a large volume of data to cloud storage to work on the files at home. Instead of malicious activity, some activity is only a minor infraction of a local policy (Balakrishnan, 2015).

The Management Reporting step (#12) involves keeping the InT hub focused on what is important to the organization. It is a means by which the leadership can monitor the progress of the analysts and maintain the integrity and relevance of the InT hub. This also allows for communications of the good progress the hub is making to help push future requests for resources.

Finally, the Feedback and Lessons Learned step (#13) provides the always-important self-learning mechanism of any organization. It examines what worked and what did not and then uses this knowledge to improve operations. This can result in modified rulesets within UAM tools to limit the number of false positives an analyst has to sift through and the amount of false reporting that may result.

Lee (2015) takes a different approach with his sliding scale of cybersecurity. His model focuses more on the large muscle movements of the people, technologies, and processes in the cybersecurity effort. Figure 4 displays the five stages of the sliding scale of cybersecurity. This scale can also represent a model of what happens in an InT hub



The active defense and intelligence portion of the sliding scale are the stages most closely related to the active analysis occurring within the hub.

Figure 4. Sliding Scale of Cybersecurity. Adapted from Lee (2015).

Lee's diagram can be used to characterize the major components of a hub at the people, technology, and process interaction level. Each stage describes what must happen in order for an InT hub to detect, analyze, and report on an insider threat action for mitigation procedures.

- The Architecture reveals the cyber systems in use at an organization with the notion that security must be a key consideration. This means the system must have mechanisms in place to allow observation of the activities of the users. This stage is similar to steps one through nine of the previous model.

- Passive Defense represents the systems that monitor cyber activities. This includes intrusion detection/prevention systems (IDPS), UAM systems, and security information and event management (SIEM)

systems, among many other possible monitoring systems. These represent the technologies that the hub can use as key contributors to the information the analyst must examine.

- The Active Defense and Intelligence activities capture the practice of using analysts to consume the information generated by the sensors and other inputs in order to detect malicious insider activity and the steps taken to mitigate that activity. This is a key stage in the success of an InT hub. Without the analyst, the data does not become information, which cannot then become intelligence.

- Finally, Offense is the culmination of the entire process, which is taking actions to mitigate the insider threat. This encompasses establishing the facts that a threat materialized and an attempt (possibly successful) to use authorized access for malicious or unintended purposes occurred (either intentionally or unintentionally). From this stage in the process, organizations may take legal or other administrative action on an individual.

Chapter IV examines this entire scale of cybersecurity operations to better scope the nature of the information-sharing problem. For the remainder of this section, however, the focus is on the intelligence process to understand how the analyst takes the data and turns it into information and intelligence to be useful in the mitigation process.

According to Lee, five steps comprise the intelligence cycle: planning and direction, collection, process and exploitation, analysis and production, and dissemination and integration. Evaluation and feedback also permeate the five processes (Lee, 2015). Figure 5 displays these processes and their relationships.

The intelligence process contains the five sub-processes represented within the inner circle. The evaluation and feedback processes occur within each of the sub-processes. Even though the processes are sequential, it is important to note how they come full circle and all end up serving as input to the other processes.

Figure 5. The Intelligence Process. Source: Lee (2015).

Intelligence is "the process of collecting data, exploiting it into information, and producing an assessment that satisfies a previously identified knowledge gap" (Lee, 2015). To go further with this definition, the parts of the intelligence cycle can apply to cybersecurity and insider threat monitoring as well. In this model, each stage is sequential; however, the results of one stage ultimately circle back around to become possible input factors for previous steps.

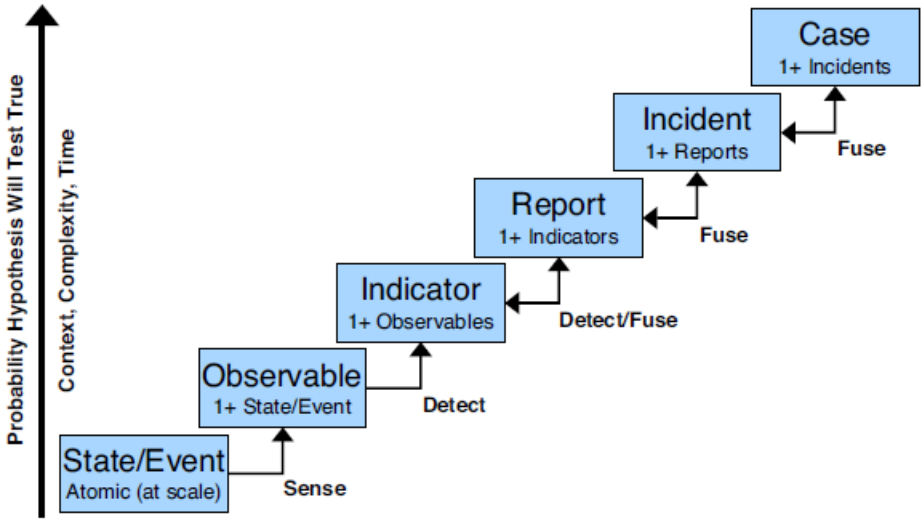
The planning and direction stage, Lee explains, encompasses setting up the network and the analytical framework of the organization as well as policy generation. The collection stage uses the tools within the system to monitor and report on user activities. In the process and exploitation stage, the UAM matches activity to rulesets and presents the relevant activities to the analyst.

Analysis and production encompasses fusion of the gathered information and its transformation into intelligence as well as producing case reports that will be elevated to

leadership. Report development is one of the most important activities of analysts, as this step cannot occur without their interaction. Tools alone cannot perform the analysis of the information, “only analysts can create intelligence” (Lee, 2015).

In the dissemination and integration step, producers give the report to consumers requiring the information, such as law enforcement or the supervisor of the threatening individual. This intelligence can then feed back into the UAM system to create thresholds and rulesets that are more refined for the identification of future threats (R. Laylo, interview with author, March 14, 2015).

The research presented concepts behind the process, but the artifacts generated are also important. The artifacts are what are processed, analyzed, and communicated among InT hub personnel and to external entities that can lead to mitigating actions. Brackney and Anderson (2004) provide a representation of the major elements that move through an InT department in Figure 6.



This figure shows the compilation of how an atomic event can grow to become a case within the insider threat hub. The horizontal axis is implicitly showing causality of the events below it within the figure. The vertical axis represents the analysis process performed on the artifacts over time.

Figure 6. Sequence of Events in an Insider Threat Case. Source: Brackney & Anderson (2004).

When one or more events or system states occur, some observable element triggers an alert. The UAM detects this element and creates an indicator. Depending on the thresholds and triggers set up in a system, this can lead to a report of the event. One or more reports can lead to an incident, which can subsequently turn into a case. Then the case is elevated within the insider threat hub until it reaches a final decision point for action or non-action.

This concludes the research coverage of the processes that go into establishing an insider threat hub, the intelligence cycle, and the sequence of events within the hub. A final technology for examination is that involved in ingesting and storing UAM data.

C. DATABASES

Another background area for the insider threat information-sharing problem relates to the organization of data within the InT hub databases. In order to be effective in sharing both indicators and case data, this research examined the databases housing the data. The UAM system receives insider threat information from multiple logs and other sources for the organizations interviewed for this research. UAM data resides in a relational database, so an understanding of how this technology works and its shortfalls are crucial to solving the problem of sharing data and information across InT hubs. In particular, the disparity between database schema and merging this information could prove to be a major challenge for data sharing that relies solely on the relational database technology.

The Relational DataBase Management System (RDBMS) dates as far back as 1970. A 1970 Communications of the ACM article, written by E. F. Codd, discussed the need for data structures that move away from hierarchical methods (much like how computers store files in folders and subfolders). Codd stated there existed a need to evolve to data structures with data independence. Data independence means that organizations can change certain aspects about the data without also having to change other, higher-level factors related to the data.

There are two main types of data independence in databases: logical and physical. Logical data independence means an organization can change the structure of the

database without having to change external programs that access the data (Elmasri & Navathe, 2000). For example, if you add a field to a table within a database, the existing programs that access the original fields should be unaffected. Alternatively, if a database administrator removes a field, all applications that only access the remaining fields should be unaffected.

Physical data independence means an organization can change the internal structure of the data without having to change the external schema (Elmasri & Navathe, 2000). An example of this is changing the access path to data. If the data is the same within the database, the external schema will not have to be altered (Elmasri & Navathe, 2000).

Relational databases provide the means to give data the independence Codd was referring to and allows for fast recall of the data. Relational databases also exhibit a strict structure. The data is stored in columns of a table. The columns represent a field where values are of a certain type, such as an integer or a string of a certain length. Values in a row link with other data in the same row. Figure 7 displays an example of a few relational tables that may be stored within a database.

Employees		
Employee_ID	First_Name	Last_Name
12345	Shaun	York
12346	Danielle	Smith
12347	Robert	Tindle
12348	Simon	Elliot
12349	Stanley	Elliot

Departments		
Dept_ID	Dept_Name	Dept_Head
N1	Administration	Danielle Smith
N2	Intelligence	Simon Elliot
N3	Operations	Robert Tindle

Badges		
Badge_ID	Employee_ID	Clearance
01234	12345	Confidential
01235	12346	Top Secret
01236	12347	Secret
01237	12348	Secret

Rooms		
Room_ID	Room_Name	Floor
01-243	Conference Room	1
02-323	Operations Center	2
02-324	Server Room	2

Figure 7. Example Database Tables

In the first table, Employees, there are three fields: Employee_ID, First_Name, and Last_Name. It is obvious in this simple example that Simon Elliot is an employee

with an employee ID number 12348. This table relates an employee (by name) to their identification number.

Relational databases operate based on relational algebra, which is rooted in set theory. An operator defines how a database combines and sorts data within a set. The operator interacts with the data via a Structured Query Language (SQL). The following SQL query is an example of how an operator can interact with the data within the tables in Figure 7.

```
SELECT *
FROM Employees
WHERE Last_Name = ‘Elliot’;
```

This command string directs the database to select all columns from the Employees table and filter just the rows where the entry in the Last_Name field is “Elliot.” Table 1 shows the result of the aforementioned query.

Table 1. SQL Query Result

Employees		
Employee_ID	**First_Name**	**Last_Name**
12348	Simon	Elliot
12349	Stanley	Elliot

As displayed, the query returns the two employees Simon and Stanley Elliot. SQL queries become more involved when they query more than one table, introducing the application of relational algebra. For example, you can run the following query, again using the tables in Figure 7:

```
SELECT Employees.First_Name, Employees.Last_Name,
Badges.Badge_ID

FROM Employees

INNER JOIN Badges

ON Employees.Employee_ID = Badges.Employee_ID

WHERE Badges.Clearance = 'Secret';
```

This directs the database to select the First_Name and Last_Name columns from the Employees table as well as the Badge_ID column from the Badges table and present a list of employees who hold a secret clearance. The dotted notation of the SELECT statement instructs the RDBMS which columns from which tables to reference. The INNER JOIN statement temporarily allows the program to combine the tables based on the ON statement, which matches on values in the Employee_ID values in the Employees and Badges tables. Table 2 shows the joining of the two tables with the requested columns presented after applying the filter of only displaying those employees who have a secret clearance.

Table 2. SQL Query Result

Employees		
First_Name	**Last_Name**	**Badge_ID**
Robert	Tindle	01236
Simon	Elliot	01237

Within the relational database structure, several main components allow this structured query to work. The components are:

- **Table Name:** This is an identifier for an entire table within the database (e.g., Employees). These must be unique within the database.
- **Field Name:** This is an identifier for the field of the table (e.g., First_Name). These names must be unique within a table.

- **Field Value:** These are the actual data values within the database fields (e.g., Robert). These values do not have to be unique, but must adhere to a predefined structure (e.g., integer, string, etc.).
- **Primary Key:** This is the column value or group of column values that makes the rows unique within a table (e.g., Employee_ID).
- **Foreign Key:** This is a primary key of another table referenced in a table, which allows for the relation of the data to be established (e.g., Employee_ID in Badges Table).
- **Cardinality:** This is a database structure rule that dictates the relationship (based on ratios) of one entity to another (e.g., an employee may only belong to one department (1:1), but a department may have many employees (1:N)).

Relational databases record transactions about information. To ensure reliability of these transactions, a relational database must adhere to ACID principles (Gray, 1981). ACID stands for atomicity, consistency, isolation, and durability. **Atomicity** means all transactions are all or nothing. For example, if someone made a funds transfer from one bank account to another, the transaction must be complete from start to finish. If a system outage occurred mid-transaction, then, without atomicity guarantees, the money could disappear and be in neither account. Until all steps to a transaction are final, they cannot commit (complete). **Consistency** means a database is always in a valid state. No transaction can result in a data value that breaks either field value rules or cardinality rules within the database structure. **Isolation** means until a transaction is committed, it must remain separated from the effects of any other transactions. Nothing can write or otherwise manipulate the data contained within the transaction. **Durability** means the system retains record of the transaction, even if there is a system failure. A database accomplishes durability by saving records of transactions to the non-volatile hard disk of a computer versus volatile Random Access Memory (RAM); this action retains the record in the case of a power loss.

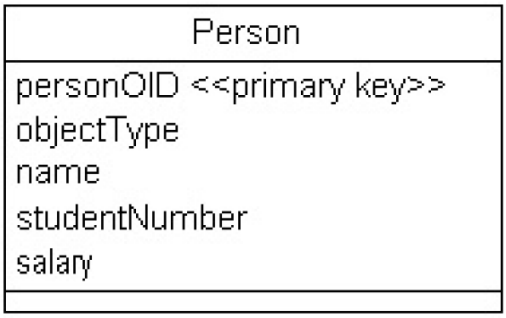
This need to adhere to the ACID principle creates a storage mechanism that allows for strong protection of data, such that changes to the data are not allowed that would place the data or the database in an invalid state (i.e., inconsistent). Also, since the

data within a relational database has a simple data type (e.g., integers, strings, etc.), it allows for quick access and manipulation of the data via compare and combine methods.

However, real-world activities, such as insider threat detection, is not something that lends itself well to simple data types. If hubs use different data types and if these differing data types reflect different semantics, then sharing information among hubs may be very challenging. Something as simple as a different format used for a date and time can make two databases unable to communicate without a programmed solution installed between the two interfaces. If something more subjective, such as a measure of loyalty to the United States is stored and shared, the meaning, i.e., semantics, of that qualitative value may be more difficult to express and share. Humans exist in a world where they interact with objects. Furthermore, the interaction with objects exhibits a type of behavior. Relational databases, based on the relational model, do not always capture object interactions or the related behavior associated with those interactions. Alternative models could be better suited to handle the task of facilitating communications between systems with differing data types and data semantics.

Ambler presents some useful examples of an “impedance mismatch” that exists when trying to capture object-based entities (objects with attributes) within a relational database (Ambler, 2000). He uses the example of students and professors represented within a relational database. The attributes of both of these types of people are within fields of a database. He introduces a unique (but often examined) problem regarding the relationship between students and professors, specifically how to represent the same person who is a student AND a professor at a university.

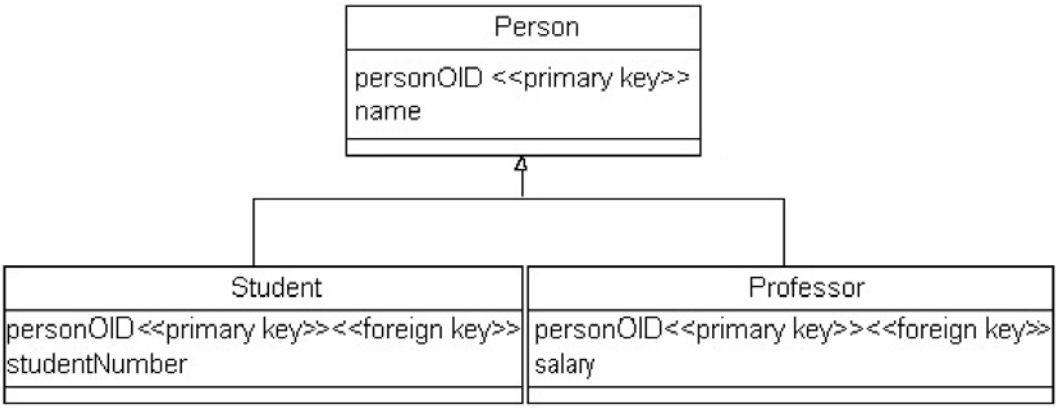
One of the challenges in representing these types of data within a relational database is that, while both types of people are organizational people, they exhibit different attributes (e.g., professors have a salary; students do not). Trying to map the same person to both a student and a professor can be cumbersome. Ambler presents several methods to represent the data within a relational database. Figure 8 displays one of his approaches.



This is an attempt to save all persons (student and professor) as one entity type within the database.

Figure 8. Saving All Persons as One Entity Type. Source: Ambler (2000).

This method strives to assign a person an object type, such as "student" or "professor." This has one clear drawback, which is representing a student who is also a professor, since the object type is now two separate values (Ambler, 2000). Ambler provides several more methods that entail saving students and professors as their own entity types, which introduces yet again the challenge of representing the same person who is a student and a professor. He provides a solution where he assigns a value called "PersonOID" which is a unique primary key to the Person type and a primary AND foreign key to the student and professor types, displayed in Figure 9.



This method attempts to create a class called "Person" and establish two sub-classes of "Student" and "Professor." The primary key of the Person table, "personOID" becomes the primary AND foreign keys for the two sub-classes of "Student" and "Professor."

Figure 9. Using Primary-Foreign Keys for Multiple Sub-Classes of Person. Source: Ambler (2000).

This method creates a different problem of populating the database with many tables, which effectively duplicates entities and slows read and write times due to having to access multiple tables (Ambler, 2000). As data grows exponentially, this slow-down and wasted space become an even larger issue. Ambler introduces other methods to try to solve this student/professor problem with relational databases, but he shows that each method has its “trade-offs” (Ambler, 2000).

This overall example is but one instance of how relational databases can fall short at even simple tasks when attempting to represent objects within an inflexible ACID-based structure. Relational databases are good at storing simple data types and creating associations via relational algebra. However, when trying to represent objects, relational databases have to take shortcuts, which can present additional burdens on the database. With an increase in the amount of data (viz. “big data”), distributed file systems, and data transfer rates, relational databases are not always the most effective means to manage the data. There needs to be a better solution.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

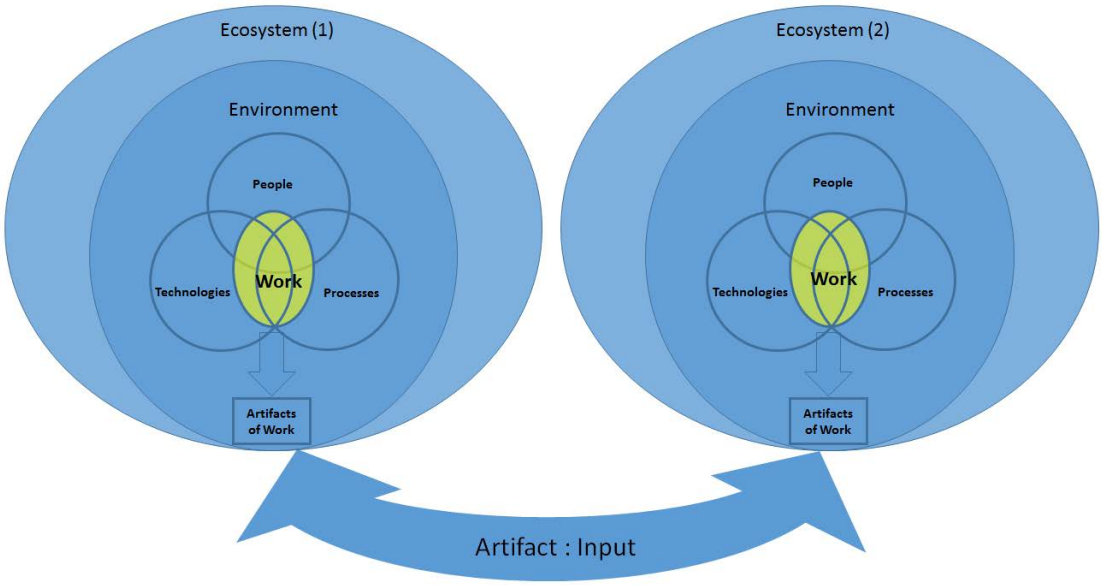
The InT hub exists to carry out the special mission of deterring, detecting, and mitigating the insider threat. The organization of an InT hub, based on the mission, must have a minimum set of resources and personnel to operate. This organization and the internal elements were of paramount importance for this research to discover how the data moves through the organization to establish a means to share the data between InT hubs. To understand the problem of data and information sharing between hubs, two steps were required. First, I framed the problem using a model. Then there was a need to examine the actual data via a structured systems analysis. The goal of both of these methodologies was to answer the research questions provided in Chapter I:

1. What is the known set of cyber methods used by insiders?
2. What workflows are necessary to InT hub operations that require integrated data within and between hubs?
3. What technical means are required in order to share information between hubs?

Chapter II(A) provided an answer to the first question. The second and third questions required research into how the InT hubs operated. Personal interviews with members of InT hubs from several organizations provided the information needed for both of these analyses.

A. THE MODEL

The InT hub is a collection of elements working in concert with one another to perform the tasks of deterring, detecting, and mitigating insider threats. There was a need for a model to provide a visual representation of the complete picture of how all the pieces fit together and to guide the research. This model needed to include the relevant pieces that dealt with the data flowing through the hub to explore the link between multiple hubs and how the data could make the transition from one to another. Figure 10 displays the derived model for this problem set.



The insider threat problem exists within an ecosystem for an organization. The ecosystem is the interaction of the people, technologies, and processes operating together within the environment of the organization. All of these components are doing work and producing artifacts of that work. The artifacts must be able to function as input for other ecosystems in order to be effective.

Figure 10. Ecosystem Approach to Insider Threat Sharing Problem

This conceptual model represents the people, technologies, and processes within an InT hub doing some form of work, resulting in the creation of artifacts. These artifacts of work must somehow become a useful input for other hub(s) within the insider threat community. It was imperative to identify the specific artifacts that would become useful to other hubs and how an intermediary solution would present those artifacts to other hubs (e.g., format, data type, etc.). Successful implementation of this research model required knowledge of the people involved with the InT hub operations, the technologies used, and the processes performed. An examination of the artifacts of the work could correlate how an intermediary solution would export them to other InT hubs. These artifacts could provide both InT indicator data as well as InT case data, thus facilitating better operations within the hub.

Starting at the macro level, the first level to explore was the ecosystem. These include: the definition of an ecosystem is any collection of connected and interacting parts; and the people, technologies, and processes are definitely connected and

interacting, performing work and creating artifacts. As discussed in Chapter II, the intelligence cycle is circular; the output of one part of the cycle feeds all other parts of the cycle, either directly or transitively through intermediate phases of the cycle. In addition, all of these interacting and interconnected parts operate within a specific environment. Whether the environment is defined to be the culture, the platform (i.e., afloat, shore-based, etc.), or necessary operating parameters, it imposes restraints and constraints to operations.

The **people** within an organization have a direct connection to technology, specifically with the cyber monitoring of employees. The people encompass not just the InT hub workers, but also the employees of the organization who are being monitored as well as the subject matter experts (SME), who may be located outside the physical hub structure. The people give the data its meaning, both in the creation of the monitored activity and the analysis of what the data means. For example, a monitoring solution captures an employee performing an action. Performance of the action has meaning.

The **technologies** within an organization exist in two sets: technologies used by employees in the performance of their jobs and the technologies used to monitor activity. In some cases, the physical technology can be the same system, just used in a different manner to meet the intended requirements. The work performed by these technologies creates data. The technology also consumes the data, so an understanding of the interactions of both sets of technology and how the people interact with them and the resultant processes is critical.

The **processes** encompass the steps involved in operations of a hub. Processes can be broken down into sub-processes until reaching an adequate level of understanding about the data types within the system in question. The processes allow a view of the flow of data throughout an organization and can assist in locating problem areas or areas of opportunities for improvements or additional functionality.

The **work** encompasses the actions performed by the people, technologies, and processes, including the interrelations between these elements. The work is purpose-driven, whereas the tasks are the detailed steps required to complete the work. This work

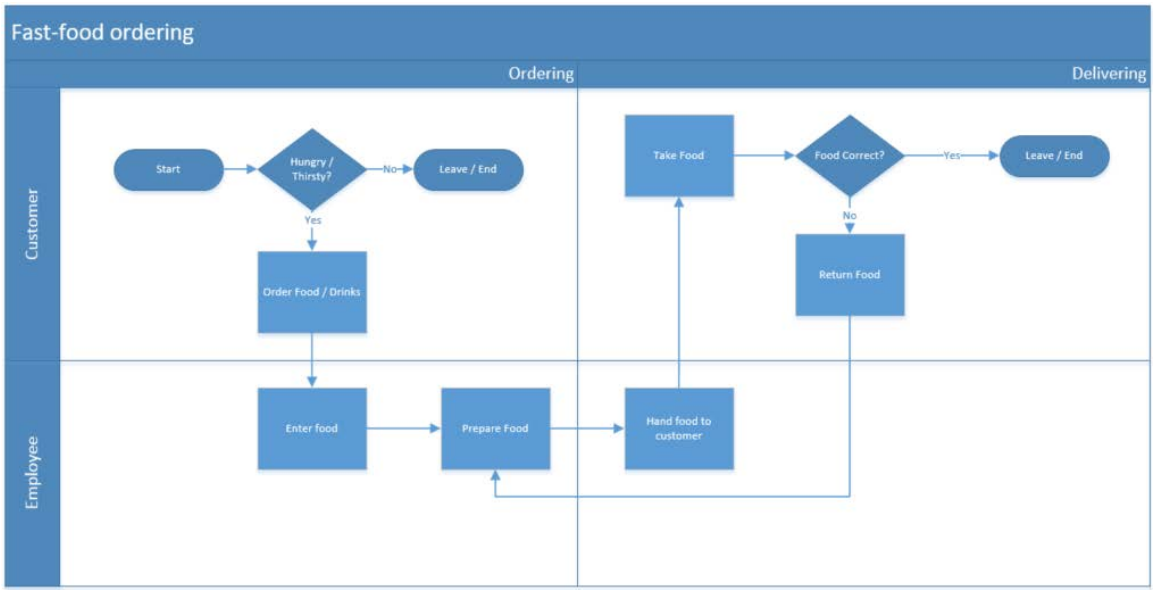
generates the artifact objects that an organization must manage in order to become useful inputs to another InT hub's operations.

The **artifacts of work** are those objects remaining after work is complete. Whether a hard-copy document from a print job or a log entry from a cyber activity, these are both artifacts that become recorded instances of work performed. A comparison of these artifacts with the necessary inputs to an insider threat hub could result in discovering how to translate from one to the other.

Finally, a **link** must exist between separate ecosystems to allow organizations to use artifacts from other organizations as input. The ecosystem of one organization may be vastly different from that of another. This difference must be a considered when developing a solution for inter-hub sharing of information and data.

B. STRUCTURED SYSTEMS ANALYSIS

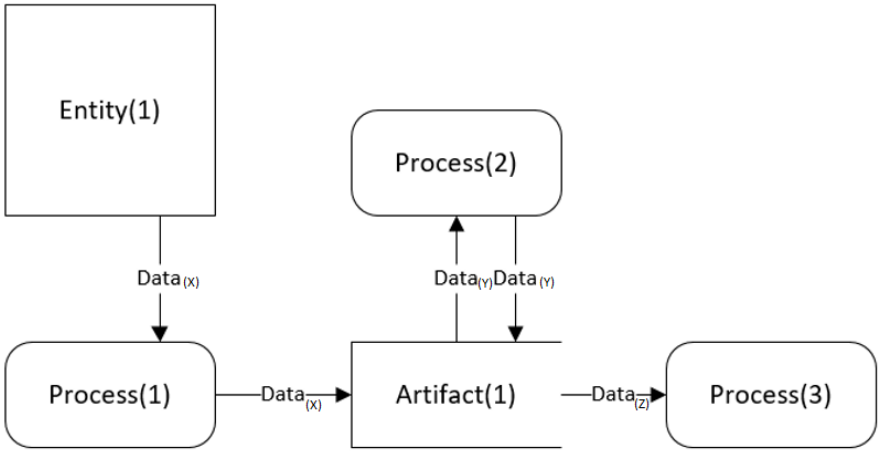
To examine the elements of an InT hub within the aforementioned model framework, I conducted a structured systems analysis methodology using workflow and data flow diagrams (DFD). This resulted in a visual representation of the flow of work and data within an InT hub that I could use to explore possible solutions for sharing data between hubs. Figure 11 is an example of a cross-functional workflow. In a cross-functional workflow, rows separate roles and columns show temporal separation.



Example of cross-functional workflow diagram using a simple fast-food ordering scenario. Ovals signify start and end conditions, rectangles represent actions, diamonds represent decisions, and the arrows show directional flow. Along the top are the phases of the operation and along the left side are the personnel involved in the operation.

Figure 11. Example Cross-Functional Workflow Diagram

The purpose of using a cross-functional workflow diagram was to examine the work performed within an InT hub. The swim lanes (horizontal rows) of the diagram provided a nice separation between the roles of personnel and the phases (vertical columns) provide a good indication of temporal flow of information. The goal was to use a cross-functional workflow diagram to capture the essence of the work and then transition to the next structured systems analysis, a data flow diagram (DFD) to capture the actual data moving through the hub. Figure 12 is an example DFD.



The Data Flow Diagram (DFD) shows an entity, three processes, an accessed artifact and the data flows between them.

Figure 12. Example DFD

In a DFD, an entity is an element outside the construct of a system that either consumes output from or provides input to a process. The processes are actions, which create artifacts that become sources for other processes. The arrow connectors show both the directional flow of data as well as the type of data that is flowing. Using this approach provided a means to learn how InT hub operations used data.

C. METHOD

In order to gain the data needed to feed the model, workflow, and data flow diagrams, I used personal interviews of members from four organizations: DOD Insider Threat Management Analysis Center (DITMAC), Defense Security Service (DSS), National Geospatial-Intelligence Agency (NGA), and the Coast Guard. DITMAC was an excellent choice due to the mission of the organization. The DOD established DITMAC to bridge the gap between DOD organizations regarding insider threat information (Defense Security Service). Conversations with leadership within the insider threat community about DOD organizations that had a good operational insider threat program identified DSS and NGA as good candidates. Even though the Coast Guard is not within the DOD, it has an operational afloat platform InT program that garnered praise from

several DOD organizations. This afloat program was highly relevant, since the Navy also has many afloat platforms.

I compiled information from the interviews and used the information to create the analysis presented in Chapter IV. Although there were some disparities within the flow of information, there were also many similarities. The interview questions and answers are in the Appendix.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS

Interviews with InT professionals from multiple organizations provided the data required for this research. The information gained filled in the knowledge gaps about how the elements of the hub work and how data flows through the system. Using this new knowledge, a more complete understanding of how hubs could share data and information now exists.

The analysis of the InT hub-sharing problem began logically at the level of the ecosystem to identify differences in operating requirements and then worked down to the sub-components of the model shown in Figure 11. My goal was to identify the work performed in an InT hub to discover how the data moved. A follow-on goal was to establish how one hub could effectively share data and information automatically with other InT hubs.

The DOD, specifically the Department of the Navy (DON), provides a broad spectrum of unique ecosystems that translate to other services and organizations. Since the Navy operates in every domain (land, air, sea, undersea space, and cyber), it is a good source for examining ecosystems as a whole. The Navy has six categories of ecosystems within the scope of this research: shore-based (domestic), shore-based (foreign), afloat, undersea, airborne, and forward-deployed. Each ecosystem has its own effect on the people, technologies, processes, work, and artifacts of work. Since the overarching mission of any InT hub is the same (i.e., deter, detect, and mitigate insider threats), the first focus for the different ecosystems was defining the environment in which each organization type operates. Within each environment, there are three main differentiating characteristics of data flow: connectivity, throughput, and expertise.

Connectivity was important because the UAM client nodes must be able to communicate with the UAM server node(s). If this communication path is not present, the artifacts of the work performed by the employees and recorded by the UAM client cannot reach the InT analyst. This means the analyst cannot analyze or report the activity. Throughput was also an issue due to the restriction on how much data the client nodes

can pass to the server node. This limitation is important in the case of full video captures of suspicious activity as it produces large files. Last, expertise was a concern due to the limitations on where an organization can support placing an InT analyst. This drives the infrastructure design and affects both connectivity and throughput requirements.

A. ENVIRONMENT

Each ecosystem has an environment that has specific parameters or attributes that differentiate it from other ecosystems. The differences can be environmental conditions (e.g., temperature, humidity, etc.), cultural, or policy-related. I discuss some differences between the ecosystems within the Navy with the focus of connectivity, throughput, and expertise.

1. Shore-Based (Domestic)

The InT hub ecosystem with perhaps the most resources available is the shore-based organization within the United States. The environment of this ecosystem allows for the greatest control over all aspects of operations, including connectivity, throughput, and expertise. Although organizations, e.g., “big Navy” and other major entities in the InT community, do not normally control the Internet service provider (ISP), each ISP within the continental U.S. has a fair selection of other competitors. This competition and a desire to maintain customers provides a significant amount of assurance for high-quality service, including connectivity.

Within the U.S., organizations normally have ready access to the latest communications technologies and mediums. This access provides organizations the ability to select, if desired, high-speed, high-capacity communications channels. An organization can choose its type of connection based on operational need and funding instead of just what scant services may be available. This means video captures of employee cyber activity can flow freely from the UAM clients to the UAM servers. Expertise is also not an issue with shore-based domestic organizations since the organization can employ both military and civilian members, based on the manning requirements. The widest selection of personnel to perform the InT analysis or supervision is available.

2. Shore-Based (Foreign)

Foreign shore-based organizations can encounter challenges in all three areas of connectivity, throughput, and expertise. On foreign soil, organizations are limited to the available ISPs provided by the host country. Depending on regulatory bodies for that country or region of the world, connectivity and throughput control could be less reliable. For example, throttling of traffic or maintenance of equipment may not be up to the standards that some domestic ISPs must maintain. This unreliability means that full video capture transmission may not be feasible. With regard to expertise, the DOD has some restrictions on personnel stationed overseas. This limits the personnel hiring pool for an InT hub at an overseas location.

3. Afloat Platforms

Afloat platforms have a distinct ecosystem created by a unique environment. When underway, far out at sea, an afloat platform has limited connectivity and throughput due to the need to use satellite communications as the primary transmission path. Although inter-vessel communications may be available, satellite communications are essential for naval vessels to communicate with shore facilities.

There are two main types of afloat vessels in the U.S. Navy when considering the ability to communicate: the flagship and the combatant ships, referred to as small-boys. The flagship of a battlegroup is normally a carrier or a large amphibious ship. Since a flagship is the hub of naval operations, there is a significant demand for absolute connectivity and exceptional throughput. Fortunately, the Navy outfits these ships to allow many communications channels to operate simultaneously. The small-boys (e.g., destroyers, cruisers, etc.), however, are not outfitted as well with communications gear. Due to equipment and channel allocation restrictions, the small-boys do not enjoy as much throughput as the flagship. In addition, due to antenna positioning and the state of the sea, the ships may experience degradation of connectivity, making it difficult to get UAM client data to the UAM server node.

For ships of either type, the pool of personnel to select from for InT protection is more limited than either of the aforementioned shore-based organizations due to issues

that can make an individual unfit for sea duty. In addition, manning restrictions, additional duties for assigned members, and a restricted ability to support on-demand riders limit how many people can be onboard a vessel. This means that having a dedicated analyst onboard a small-boy may not be feasible.

In a March 30, 2016 interview with Jared Quance, the Assistant Director for Insider Threat U.S. Coast Guard Counterintelligence Service, he provided a possible solution for this afloat challenge of not being able to station an InT analyst onboard small-boys. He stated it might be possible to remedy the connectivity, throughput, and expertise problem by stationing a group of InT analysts onboard the flagship and to allow them to analyze the activity for all of the members on all of the ships of the battlegroup. This would mean the smaller ships could connect to the flagship to download the UAM client information to the UAM server as connectivity and bandwidth allowed. He also explained how Coast Guard successfully queued up UAM data on afloat platforms as low priority traffic, transmitting the data as bandwidth became available. This queuing method proved the reliability of UAM data transmission while at sea without significantly affecting the bandwidth used for mission-essential communications. Having the small-boys send UAM data to the flagship would also alleviate the manning limitations on the smaller ships, since support for on-demand riders is better on a flagship.

4. Undersea

Undersea platforms are subject to even greater restrictions than the afloat platforms due to restrictions on the former to communicate resulting from the covertness of their operations. Their opportunities to surface in order to communicate with satellites or other units can be limited. This limitation means their ability to transfer UAM data to an analytical cell may only occur when they tie up to a port and connect to a hard line communication channel. This would create a significant delay in the detection process, which, transitively, delays all the subsequent steps as well. This may be an inherent risk the InT community must accept. Undersea platforms have even more stringent requirements for their members; medical and family conditions, as well as security

clearance restrictions, remove even more people from pool of possible members. Undersea platforms may not have the space for a dedicated InT analyst. In addition, the undersea vessel would have to maintain enough storage to store all of the UAM data until it was able to transmit the data to the UAM server once pier-side.

5. Airborne

Airborne platforms create yet another challenge based on connectivity, but are also affected by possible covertness of operations. Although airborne assets are likely to have better line of sight for communications purposes, they may not have the extensive communication equipment necessary to transmit UAM data while in flight. In addition, due to the stringent restrictions of available space and job role requirements on an aircraft, this would make it difficult to perform the InT analysis onboard. However, an aircraft must land and refuel regularly, which serves to benefit the UAM data challenge. This allows opportunities for an airborne platform to plug into a landline or connect wirelessly to a ship or land-based host to download any UAM data that may be resident on the aircraft. This may be more relevant to larger aircraft than smaller craft with no user-relevant data processing equipment onboard. A final challenge for airborne platforms is that they may be parked for long periods without connecting to a UAM server. An InT hub organization would have to factor this into the InT plan.

6. Forward-Deployed

A forward-deployed organization, in this context, is a unit that does not have the normal cyber facilities associated with a shore-based organization. The unit may have limitations on connectivity due to the covertness of its operations or equipment limitations. A common method for forward-deployed units to communicate with satellites is via mobile communications platforms, such as a very small aperture terminal (VSAT). This is a communications channel that may be limited in throughput and even connectivity based on the geographic location of the unit. In addition, due to limited supplies of forward-deployed units, there are likely to be limitations on the number of personnel they can support. This means that it is unlikely that a forward-deployed unit can support a dedicated InT analyst on site. This situation would require InT data be held

on station and relayed later (for covert operations) or to low-priority queueing of InT data, much like that discussed for afloat platforms.

This concludes the environmental examination of the diverse platforms within the U.S. Navy. There were some significant differences within the given environments based on connectivity, throughput, and expertise. The next topic is people of the InT hub, with a focus on their roles, the work they perform, and the artifacts of work they generate.

B. PEOPLE

The importance of examining the people is to establish roles and responsibilities that uncover work tasks. The work tasks provide insight into the artifacts of that work, which further allow examination of how InT hubs can share these artifacts.

At the highest level, the organizations that comprise the core of the insider threat mission within the DOD are DITMAC and the InT hubs. The DOD created DITMAC to bridge the gap between DOD InT hubs to increase the likelihood of detecting insider threats across the enterprise by establishing insider threat reporting standards and thresholds (M. Buckley, interview with author, March 17, 2016). These standards reflect the information that InT hubs must report to DITMAC. By sharing data and information among InT hubs and DITMAC, it is possible to be better poised to deter, detect, and mitigate the insider threat.

Based on a combination of interviews and discussions with multiple members of the InT community, there are six primary groups of people involved in an InT hub: the monitored employees, information technology (IT) support personnel, InT analysts, supervisors, SMEs, and the senior official. These roles comprise the core human elements of the InT hub that allow for identification, reporting, coordination, and mitigation of the insider threat. The structure of one organization's InT hub may be different from others, but these personnel capture the main expertise elements within a generic InT hub.

1. IT Support Personnel

Information technology (IT) support personnel are people with several different skillsets and roles. Included in this group are the database administrators, system

engineers, system administrators, and any other person that is responsible for how data is stored, coded, or transmitted within the electronic systems of an organization. Although the tasks performed by these personnel are vast, the two most important tasks related to InT hubs are ruleset modification and database management.

The rulesets are the thresholds coded into a UAM tool which, when matched against activity on the network, provide a method to identify and prioritize possible malicious or dangerous activity. Ruleset creation can be more of an art form than a science. For example, a rule may dictate that the UAM flag any email traffic containing the word “stab” as a possible malicious indicator. However, if an employee sends an email to someone describing how she is going to "take a stab at completing the new project training course," this is not an indicator that someone has violent tendencies or malice toward her coworkers. An organization can alter rulesets over time if they result in many false positives (benign behavior that the automated system flags as suspicious). The InT supervisor approves any changes to the rulesets and sends them to the IT support personnel for application. To use the previous example to show how an organization can modify a rule, the new rule could be to match on all mentions of "stab" within emails unless within two words of “completing.” This could alleviate many false positives without creating a gaping void for false negatives (malicious activity categorized as innocent behavior) to slip past the automated tools. To achieve checks and balances, the personnel who update the rulesets within the InT hub should be separate from those doing the InT analysis. This is a key security feature to ensure the analysts are not creating rulesets that discriminate against specific people (R. Laylo, interview with author, March 14, 2015).

The IT support personnel also maintain the security and structure of the databases used for InT analysis. For DSS, the UAM tool stored all activity in a relational database. This data must be structured and secured in accordance with federal and organizational rules and regulations. The database administrators must make any changes to the structure or security of the database. As stated in Chapter II, the database structure of a relational database is rigid and must adhere to strict ACID principles. This rigidity can

result in significant delays between a needed change and its actual implementation. Table 3 shows the work and artifacts of work performed by the IT support personnel.

Table 3. Work and Artifacts of Work for IT Support Personnel

Work	**Artifact of Work**
Update ruleset	Ruleset data
Change database configuration	Database configuration data

2. INT Analyst

InT hub analysts are perhaps the most important members of the InT hub architecture. They are crucial to analyzing the data provided by the UAM, adding context and critical thinking to the recorded activity, and creating recommendations based on their analyses. The InT analyst is skilled in critical thinking techniques and methods to find malicious activity from a series of seemingly innocuous activities. They are also knowledgeable about the many nuances of the insider threat, the UAM tool, the analytic process, and are aware of the legal restrictions under which they operate. Table 4 lists the work commonly performed by InT analysts and the artifacts created because of that work.

Table 4. Work and Artifacts of Work for InT Analyst

Work	**Artifact of Work**
Analyze indicators	Analyst decision
Fuse indicator data with other data	Modified indicator (with comments)
Comment on indicators	Modified indicator (with comments)
Clear indicators	Modified indicator with a cleared flag
Upgrade or downgrade indicators	Modified indicator with new priority flag and comments
Elevate indicators to supervisor	Modified indicator (with comments)
Request enhanced monitoring	indicator data (with comments)
Request SME support	Modified indicator (with comments)
Make a case	Case data
Update a case	Modified case data
Request ruleset update	Draft ruleset

For the analysis task, the artifact is a decision the analyst makes to move forward or not with their analysis of the indicator. While this is not a physical artifact, this is important to follow-on tasks. The fusing of the indicator data with data from other sources produces an artifact of a better understanding of the nature of the indicator. During this process, the analyst makes comments about the indicator data. This work covers the artifact for the two activities of fusing data and commenting on the indicator listed in Table 4. If the analyst clears the indicator, this archives the event and removes it from the list of alerts for analysts to review, but maintains the event within the system.

The act of an analyst upgrading or downgrading an indicator creates an artifact of a modified indicator with a new flag or assigned priority. If the analyst elevates the indicator to the supervisor, this creates an artifact of a modified indicator as well, since the analyst makes notes before forwarding. If the analyst requests SME support, the analyst sends an indicator to the supervisor, who makes the decision whether to pass it on to a SME. If the supervisor elects to pass it to a SME, he first directs the analyst to make the indicator a case, which is the next task that creates an artifact of a case file. This case file contains indicator data as well as other supporting data, but ultimately, the artifact is a case file. When the analyst updates a case file, this creates an artifact of a modified case file.

When an analyst notices many false positives on a certain ruleset, he has the ability to request a ruleset addition, removal, or modification within the UAM tool. The supervisor approves this request, but the artifact of the analyst's part in the ruleset change is a draft ruleset.

3. INT Supervisor

The InT hub supervisor is another important member for InT operations. This person provides both oversight and guidance to the analysts and can serve as a liaison between the analyst and external SMEs. In some organizations, the analyst must communicate through the supervisor to get information to or from the SMEs, but other organizations choose to provide a means by which the analyst communicates directly with the SME and carbon copies the supervisor. Both models have benefits and

drawbacks. For the analyst-to-SME direct link, the time to receive an answer can be minimal because there is no middle person adding time and effort to the process. However, if the supervisor is in the flow, he can mitigate some of the unnecessary information requests and be able to answer some of the questions presented. Section 2 discussed other supervisor activities. Table 5 provides the work performed by the InT supervisor and the artifacts of that work.

Table 5. Work and Artifacts of Work for InT Supervisor

Work	**Artifact of Work**
Review Indicators	Supervisor decision
Comment on indicators	Modified indicator (with comments)
Direct to clear indicators	Message to analyst to clear indicator
Direct to upgrade indicators	Message to analyst to upgrade indicator
Direct to downgrade indicators	Message to analyst to downgrade indicator
Direct to request SME support	Message to analyst to make a case
Direct to make a case	Message to analyst to make a case
Review Case	Supervisor decision
Comment on Case	Modified case data (with comments)
Deny case	Modified case data (with comments)
Approve case	Modified case data (with comments)
Forward case to SME	Case data
Forward case to Senior Official	Case data
Forward case to Chain of Command	Case data
Direct to initiate enhanced monitoring	Message to analyst to initiate enhanced monitoring

4. Subject Matter Expert (SME)

SMEs are specialists who provide an expert opinion on an event, series of events, or on a person based on their domain knowledge and access to sensitive data that may not be available to the InT analyst. Depending on the InT hub architecture, the organization may include the SMEs within the hub or use them as a reach-back support if need be. These alternatives depend on how their role integrates into the hub operations. Based on whether they are included in the hub architecture, this can affect how the SME can share data or have InT hub personnel share data with them. The work performed by the SME and the respective artifacts are presented in Table 6.

Table 6. Work and Artifacts of Work for SME

Work	Artifact of Work
Review case file	SME decision
Comment on case file	Modified case data (with comments)
Access databases	Improved knowledge on subject
Return case file	Case data

5. Senior Official

The insider threat senior official is the person who has a designation in writing, to "be charged with overseeing classified information sharing and safeguarding efforts for the agency" (Executive Order No. 13587, 2011). This person may also be responsible for the final approval of involving external agencies in a case. These might include the FBI or other law enforcement agencies, depending on the organization. This person takes absolute responsibility for operations under their jurisdiction. Table 7 shows the work and artifacts of work for the senior official.

Table 7. Work and Artifacts of Work for Senior Official

Work	Artifact of Work
Review Case	Senior official decision
Approve case	Modified case data (with comments)
Release case	FBI Counter Intelligence (CI) referral
Deny case	Modified case data (with comments)

C. TECHNOLOGY

Another category of entity within the InT hub that does work is technology. The primary technologies in insider threat detection are the network user nodes, UAM clients, UAM servers, and case management system. These systems work together to monitor and analyze insider threat activity, as well as provide a mechanism to communicate and report cases.

1. Network Nodes

The network user node group includes client computers, servers, printers, faxes, scanners, and badge scanners, among many other machines that have a presence on the network and used by employees of an organization. Some of these network user nodes allow employees access to the organization's data, including employee personal data in some cases (the reason servers are contained in this “user” node list). Other network user nodes are access points to gain access elsewhere within the network. These network nodes are a needed commodity within an organization as they allow employees to do their work, but also provide a means by which to monitor activity within the organization. These devices also provide employees with an avenue to exploit weaknesses in an organization’s infrastructure to be able to modify, exfiltrate, or destroy data or systems. As authorized users, employees have the ability to work within the trust system of the network, thereby avoiding some of the scrutiny IDS/IPS provide for connections coming in from outside the network.

These network nodes also serve as an opportune location to either place monitoring sensors or pull data from system logs to feed a monitoring system to provide the initial indicators of malicious activity. Proper configuration of a UAM allows analysts to view what is happening on the network within their organization. UAM clients are the part of the UAM solution that resides on some of these network user nodes. The work performed by these diverse devices is not relevant to this research, only the knowledge that UAM clients run on some of them. The next section covers the actual work and artifacts of work.

2. UAM Clients

During the course of this research, some organizations stated that any association of an organization with a specific UAM tool could be classified. Because of this, the research identifies no particular UAM system within this paper, nor does it mention an organization's use of a specific UAM system. The UAM capabilities and operating parameters noted in this section encompass those that should be present in any industry-

standard UAM system. There may be differences among UAM solutions, but my research only mentions them in a general sense within this section.

The UAM clients, located on network user nodes, provide the first-level defense of the UAM system. These are the software components that monitor user activity at the point of network access. By placing the monitoring solution at the user edge of the network, an organization can get the best understanding of employee activity. A host-based monitoring solution can allow viewing of encrypted traffic, which provides for a larger collection of cyber observables for the InT hub.

Table 8 displays the work performed by a UAM client and the artifacts of that work. The artifacts of the work involved at the level of the UAM client are one of three things: a ruleset, an event log, or an indicator. Concerning the data moving through the InT hub system, these three data types are important. However, regarding inter-hub data sharing, the event log data is not as important due to the UAM existing between the event log and the path to another hub on the far end of the process.

Table 8. Work and Artifacts of Work for UAM Clients

Work	**Artifact of Work**
Process new ruleset from UAM server	Modified ruleset
Monitor user activity	Event log
Compare event to collection criteria	Flagged indicator data
Compare event to rulesets	Flagged and prioritized indicator data
Send indicator to server	Indicator data

The UAM clients operate based on a ruleset matching concept. Whether the ruleset includes atomic actions that are malicious (e.g., sending an email to a foreign domain) or more behavior-based activities (e.g., executing sequential processes), the rulesets govern what is reported by the UAM client. Organizations design their rulesets to minimize the false positives and false negatives. Because the UAM is trying to capture human behavior, the creation of these rulesets can be more of an art than a science. An attempt to detect every nuance of suspicious behavior with all possibilities would create a ruleset that would almost have to be infinite; this is not feasible. The rulesets are created

and then modified over time to reflect both observed patterns in false positives and engineered hypotheses of possible false negatives.

The UAM server provides the ruleset to the client. The method and frequency of the distribution of any updates from the server to the clients will be different based on the UAM vendor, but the distribution must be present. Inconsistency of monitoring on a network due to mismatched rulesets would lead to unreliable results. Once the ruleset is in place and an activity occurs, the computer logs the event via whatever logging means it has, but the UAM also records that activity. It then compares that activity to the ruleset created by the UAM software.

There are two main categories within a ruleset: general collection criteria and prioritized collection criteria. The difference between these two categories is how the UAM system presents the indicator to the analyst after the UAM server receives the data. If the UAM system flags an indicator as general collection criteria, the system places it in a low priority status in the list of indicators for the analyst to review. By having prioritized collection criteria, the UAM system can do some automated pre-processing of the data. This pre-processing can save the analyst time by giving a high priority status to the indicators that most reflect probable malicious acts and placing them at the top of the list. DSS provided an estimate of 700–1000 indicators coming in almost every week to the analyst with a review time ranging from less than a minute to about five minutes, depending on the type of indicator (e.g., text match, full video capture, etc.) (R. Laylo, interview with author, March 14, 2015). Assuming 700 incidents per week and a conservative average of two minutes per indicator, this would be at least four and a half staff-hours' worth of work every day. This is a lot of information for an analyst or group of analysts to process, highlighting the importance of a good automated pre-processing method. The artifact from this ruleset matching is a prioritized indicator.

Once the UAM system matches the activity to a collection criterion, it sends the indicator to the UAM servers. As discussed in the ecosystem section of this chapter, connectivity between the client and server node is critical. Without that connection, the analyst will never see the data. The artifact of this transmission is the indicator data in its raw format (i.e., no additional analyst comments included).

3. UAM Servers

The User Activity Monitoring servers support the next step in the UAM solution after the UAM clients send their indicator data. The servers provide the user interface to the analyst, and possibly the supervisor, to facilitate viewing and manipulating the indicator data. The servers also modify rulesets and distribute them to the clients. Table 9 shows the work and artifacts of work created by the UAM server.

Table 9. Work and Artifacts of Work for UAM Servers

Work	Artifact of Work
Prioritize and present indicators	List of indicators
Modify priority of indicators	Modified indicator
Attach comments to indicator	Modified indicator
Clear indicator	Modified indicator
Ruleset change	Modified ruleset
Transmit ruleset to UAM clients	Modified ruleset
Create a case	Case data (or indicator data)

Even though the analyst, supervisor, or IT support personnel drive the actions of the UAM server, it is included here to show the work of the server. The analyst would be responsible for the modifications to an indicator, clearing an indicator, or creating a case. The supervisor also has a role in modifying the indicator by adding comments. The IT support personnel (in the DSS case) are responsible for making ruleset changes (R. Laylo, interview with author, March 14, 2015). The artifacts of work from the UAM server encompass indicators and possibly case files. Either the UAM system or a separate case management system creates the case files. In either case, the artifacts would be the same.

4. Case Management

The case management system is the software solution for tracking evidence of suspicious or malicious activity between an InT hub and external entities. The case management system includes case files, which contain indicator information (R. Laylo, interview with author, March 14, 2015). The UAM system or a separate software suite

may be able to accomplish the case management process. The artifact of work from this case management tool is the same either way, which is a case file containing indicator data. Table 10 shows the work and artifacts of work for the case management system.

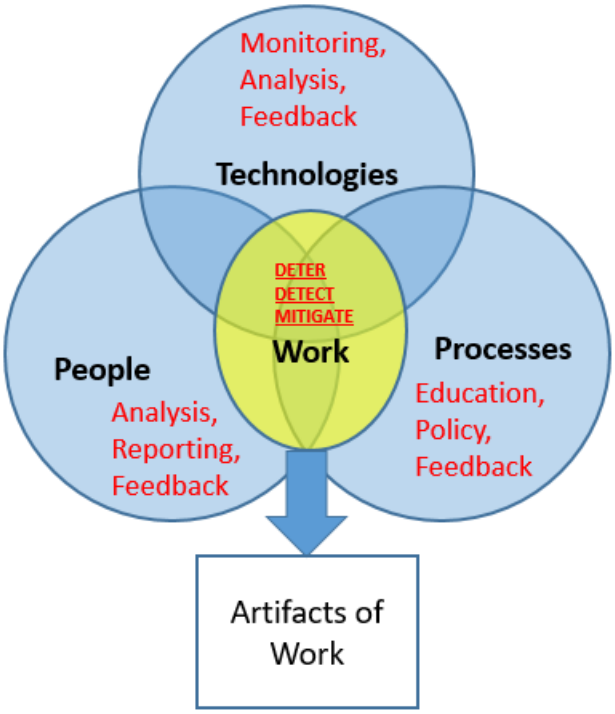
Table 10. Work and Artifacts of Work for Case Management Tool

Work	**Artifact of Work**
Create case	Case data
Modify case	Case data
Release case	Case data (or FBI CI referral)

This concludes the examination into the technologies of the InT hub. The artifacts of work discovered are indicator data, case data, and ruleset data. Depending on the situation, there may also be an FBI CI referral, but this research categorizes this referral as case data as well since the data contained would be similar.

D. PROCESSES

The final element of the work triad is the processes that occur within the hub. The primary processes of an InT hub are to deter, detect, and mitigate insider threat activity. This is universal among InT organizations. The people and the technologies perform these processes, in part, but there are also elements of the InT hub process that are outside the people or the technologies. There are also overlaps between the triad of the InT hub elements, as shown in Figure 13, which reflects a small part of the overall model for this research.



This diagram shows the three elements of an InT hub that perform work: people, technologies, and processes. Each element results in artifacts of work. In red are the contributions to the mission of deterring, detecting, and mitigating insider threat activity.

Figure 13. Triad of Work Produced within an InT Hub

The processes of an InT hub include the processes performed by the people and the technologies, but they also include work-generating elements outside the scope of the other two. The processes are all of the elements that make an InT hub operate. Education of the employees and InT hub personnel is crucial to inform everyone of what right and wrong looks like from an authorized user's perspective. Frequent and useful training of analysts can help them better mitigate insider threat activity before it even manifests in some cases. Analysts can start to notice patterns that allow them to elevate activity to the supervisor, which he can then use to notify the Chain of Command (CoC). This notification allows the CoC to intervene with an employee who may be having issues at work before an incident occurs (R. Laylo, interview with author, March 14, 2015). Policy is also important, as it is what gives the InT hub the authority to monitor its employees. It also dictates the methods of performance and management of these monitoring actions to

ensure employees' safety and privacy. Feedback is the third main process within the InT hub that must exist within all three domains of the work triad. However, within the scope of the processes, the feedback will allow the education and policy of the hub to evolve with the needs of the organization. The artifacts of work from the processes of the hub, displayed in Table 11, are training material, educational resources, and policy documents that protect its employees while allowing for the InT hub to perform its mission.

Table 11. Work and Artifacts of Work for Processes

Work	**Artifact of Work**
Education	Educational resources; greater awareness
Training	Training material; well-trained employees
Policy	Policy doctrine
Feedback	Improvement mechanisms

E. WORKFLOW

A cross-functional flowchart supplemental to this research is available at Naval Postgraduate School (NPS). Because of its size, the flowchart is presented in pieces in this chapter and illustrates the workflow of the InT hub, separated into four phases:

- recording/automated analysis
- manual analysis
- case management
- administrative/legal action

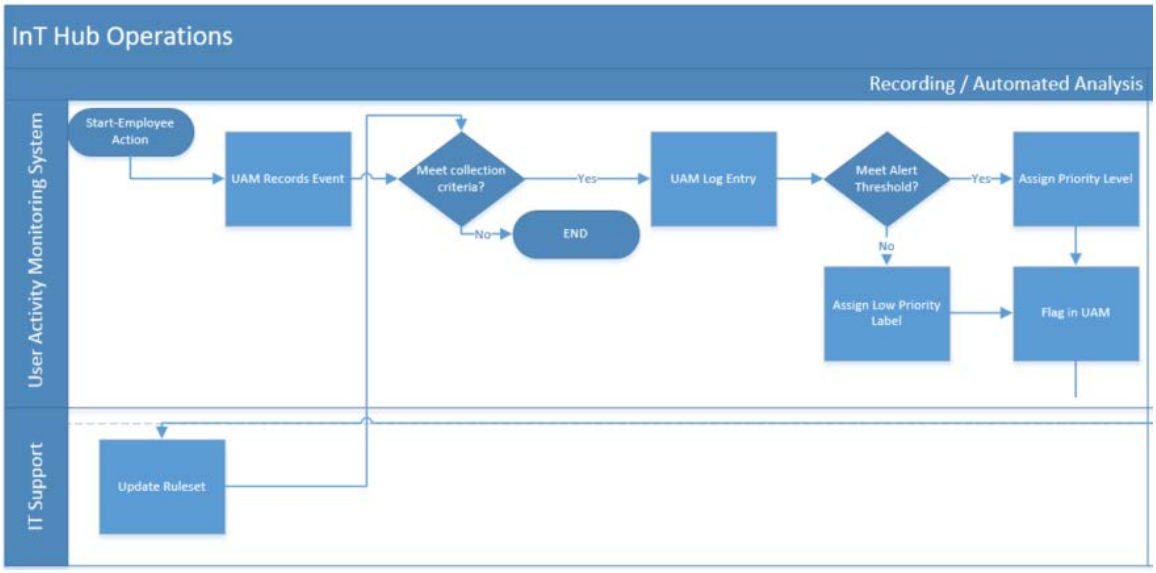
The flowchart divides the functional roles into six categories: into six roles:

- user activity monitoring system
- analyst
- supervisor
- SME
- senior official
- IT support personnel

This diagram shows the work performed within an InT hub. It includes all of the people, decisions, and processes that are accomplished. The ultimate goal of this diagram is to identify the work in order to elicit the data types that are required for InT operations.

1. Recording / Automated Analysis Phase

Figure 14 shows the initial portion of the workflow, which is an employee performing an action that a UAM system is monitoring. The UAM tool then compares this action, herein referred to as an event, to the rulesets created within the UAM tool. If there is a match against the UAM ruleset, the tool flags the event for review by the analyst. If the event matches more stringent criteria for other, higher-level rulesets, the tool will flag the event with the appropriate priority or categorization level and present it to the analyst for review.



This workflow segment shows the initial portion of the process. It includes the employee performing an action, the recording of the event, and the automated comparison performed by the UAM system. The bottom also shows the IT support personnel updating the rulesets based on input from the supervisor. The updated ruleset then feeds back into the UAM system for matching of events. Work flows to the activities shown in the workflow diagram discussed in Figure 15.

Figure 14. Initial Recording / Automated Analysis Phase of InT Hub Workflow

If the hub receives an alert from another channel, such as an email from Human Resources (HR), the organization will follow its established policy, but it normally involves an analyst opening a case or requesting to place an individual under enhanced monitoring. Enhanced monitoring means that the UAM scrutinizes an individual's cyber observable behavior more deeply, for example, by using full video capture of computer interactions. In the case of DSS, the supervisor made the decision for enhanced monitoring (R. Laylo, interview with author, March 14, 2015).

The IT support personnel play a part in this role due to the ruleset modification. If the supervisor directs modification of the ruleset, the IT support personnel make the change. The change then propagates back to the UAM clients and the clients use this new information to match events to try to discover insider threat activity.

2. Manual Analysis Phase

Figure 15 shows the next phase of the InT hub workflow, which encompasses the manual analysis of the indicator and initial steps leading to case creation. The case creation mentioned in the previous section was to show differences in how an analyst handles an indicator from different sources. This section resumes at the point when an analyst receives an indicator from the normal monitoring sources. The analyst analyzes the indicator and decides whether it is suspicious. If it is not, the analyst makes a comment and clears the indicator. He then decides if it was a false positive requiring a ruleset change to reduce the number of future false positives. If this is the case, then the analyst requests a ruleset change from the supervisor who either approves or denies the request and forwards it on to the IT support personnel. If the indicator does not require a ruleset change, then the process for this indicator stops and the analyst reviews the next indicator in his list.

If the analyst believes the indicator is suspicious, he can make one of three choices: he can request SME support from the supervisor, elevate the indicator to the supervisor for review, or he can request from the supervisor to place the individual under enhanced monitoring. Regardless of the choice made, the next person in the workflow is the supervisor.

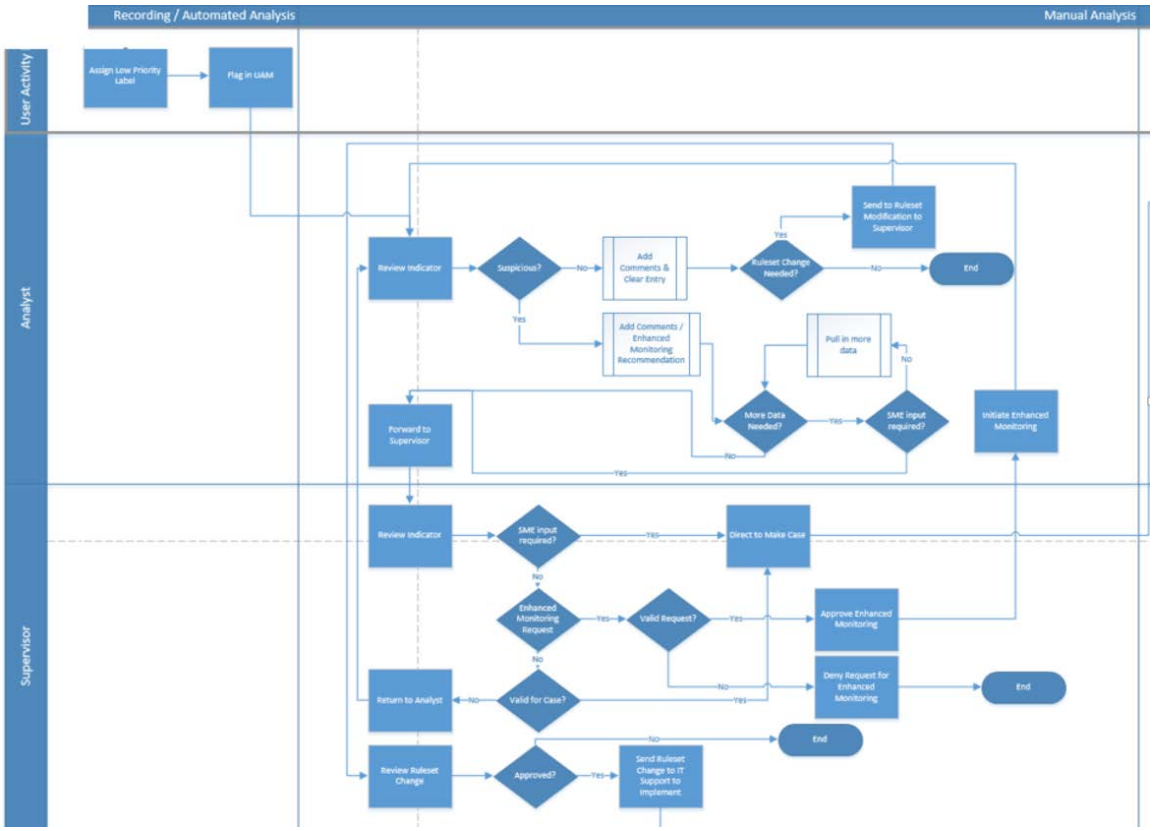
If the analyst requests SME support from the supervisor, the supervisor will direct the analyst to make the indicator a case. Any time the data leads to a report that elevates above the supervisor, the information must be a case instead of an indicator (R. Laylo, interview with author, March 14, 2015). Every case will contain one or more indicators that an individual triggered on as well as any information the analyst was able to fuse from other sources. The analyst will add any relevant comments and then forward the case to the supervisor, at which point the supervisor will forward on the case to the SME. If the supervisor disagrees with the SME request, he will return it to the analyst with instructions regarding additional information required before he will reconsider consultation with the SME.

After the SME has provided input and returned it to the analyst, the analyst will compile all the relevant information and send the case file to the supervisor with a recommendation. The movement of the case file may be a matter of marking the case file entry within the case management tool so that the case is accessible to another person. This could result in notification of the next person in the workflow that the file is ready for her review. This assumes both members have access to the same case management system. A case file can take two main paths if found to be malicious. One path is to the CoC of the individual in question, allowing his supervisor to handle the matter in-house. The other path leads either to a counter-intelligence referral to the Federal Bureau of Investigations (FBI), if an individual had provided classified information in an unauthorized manner to a foreign entity (50 U.S.C. § 3381), or to other law enforcement agencies, if illegal activity occurred.

If the analyst elevates an indicator to the supervisor who agrees it looks suspicious, the supervisor will direct the analyst to add more information if possible, and turn the indicator into a case. The analyst will then fuse data from other sources, if available and send the case back to the supervisor. The supervisor then sends the case along one of the paths mentioned in the previous paragraph.

If the analyst thinks more information is necessary to validate a suspected indicator, he can ask the supervisor to place the individual under enhanced monitoring. The supervisor can either approve or deny this request. Enhanced monitoring is

applicable if an ongoing investigation requires more information to build a case against an individual already under investigation.



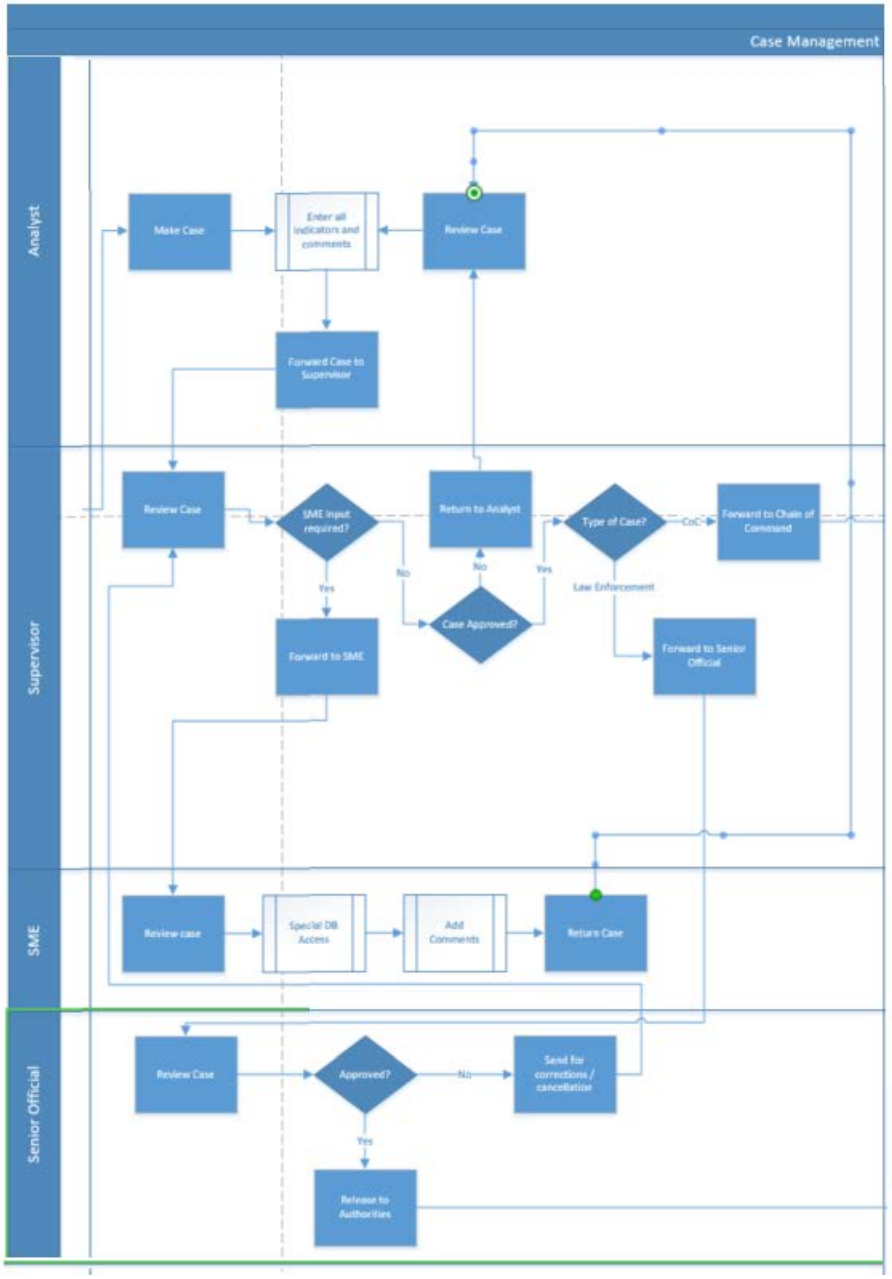
This diagram shows the workflow required for an indicator to undergo the manual analysis process. The analyst and the supervisor are the sole actors in this phase. Analysis, fusing of information, and possible case creation occur during this phase. Work flows to the activities shown in the workflow diagram discussed in Figure 16.

Figure 15. Workflow of Manual Analysis

3. Case Management Phase

Figure 16 shows the steps involved in the case management phase. Once the supervisor directs the analyst to make a case, the analyst combines all information he was able to find regarding the incident and enters it into the case management system. In the case of DSS, any time the indicators need to be elevated beyond the supervisor, the analyst must transform the indicator into a case (R. Laylo, interview with author, March 14, 2015). The supervisor then directs the case to its correct destination.

If SME input is required to get a better understanding of the incident or the individual in question, the supervisor forwards the case to the SME. Once the SME receives the case file, she will use both her expert domain knowledge in an area, such as psychology or security, and special database accesses, which the analyst does not have access to, to provide a sound recommendation. She will then return the case file to the analyst. The combination of the analyst, the supervisor, and one or more SMEs reviewing a case is sufficient for a sound decision regarding certain cases.

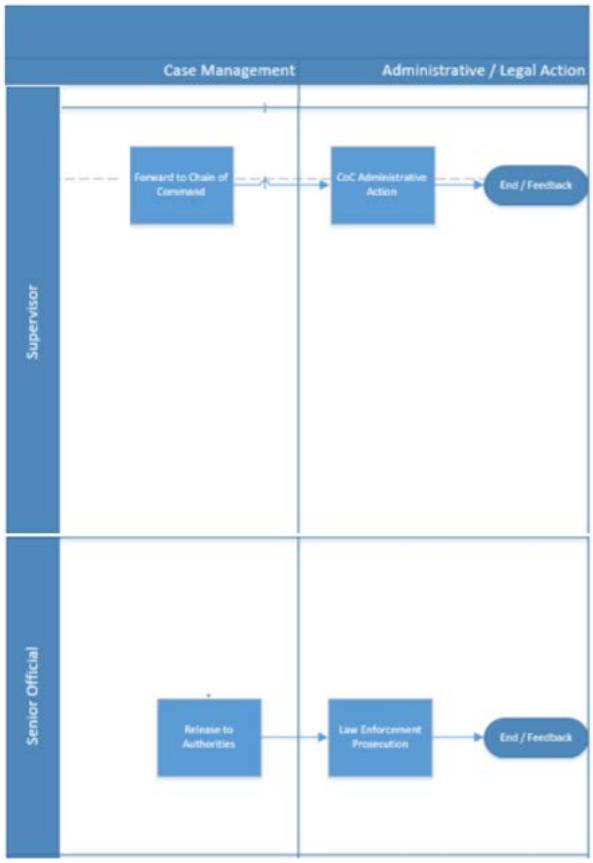


This phase starts when the analyst makes a case from one or more indicators. The analyst sends the case to the supervisor who, if he approves, will forward it either on to a SME, if required, the individual's Chain of Command (CoC), or the senior official if the matter requires release to authorities. Work flows to the activities shown in the workflow diagram discussed in Figure 17.

Figure 16. Case Management Phase

4. Administrative / Legal Action Phase

This is the culminating phase of an InT hub, during which the activities of a hub result in the mitigation of an insider threat. Administrative action may include getting the employee psychological help if he is in distress, alerting the CoC to malicious acts, or notifying the authorities of illegal activities. For DSS, if a case required administrative action, the supervisor could release it directly to the CoC. However, if the action required release to another organization, such as authorities, the senior official was the releasing authority (R. Laylo, interview with author, March 14, 2015). Once the organization takes action against the member, the ultimate goal is to get feedback from the event (e.g., did the member admit guilt, which would indicate successful InT monitoring mechanisms). The InT hub could use this feedback to modify rulesets or leverage data sources in a new way. Figure 17 shows a simple representation of the final phase of the InT hub operations.



Final phase of the InT hub operations. The supervisor releases to the CoC and the senior official releases cases to external authorities.

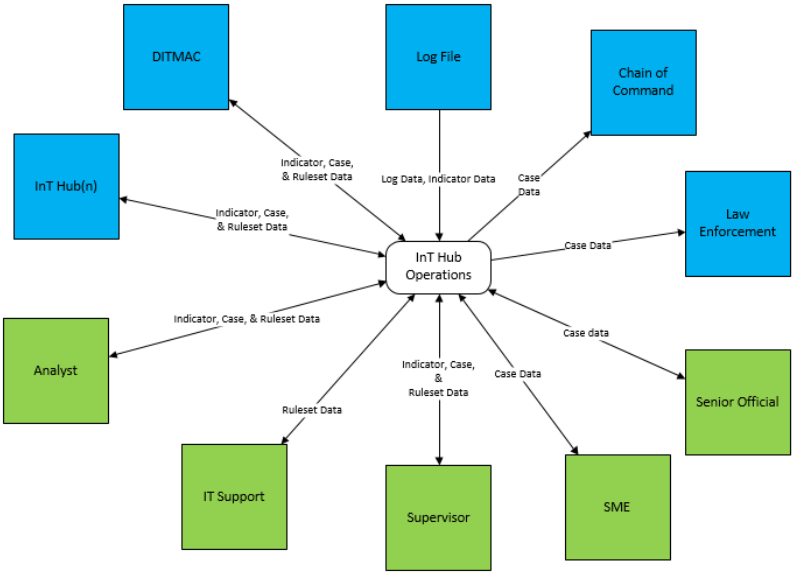
Figure 17. Administrative / Legal Action Phase

F. DFD

Based on data provided within the workflow, this showed how the organizations and personnel link within the InT community. Along these links, data travels from point to point. A data flow diagram (DFD) represents this data within this section.

Using the DFD method of structured systems analysis, Figure 18 displays a level 0 DFD, which shows the InT hub in the center with all of the entities involved on the outside. InT Hub(n) represents other InT hubs, where 'n' represents any number of other hubs. The “Log File” entity captures the concept of the employee, since a log file feeds the InT hub UAM system. This log file could be from a badge swipe system, a user terminal system, or a server. The log data types are diverse and the UAM solution configures how to interpret the data.

The green entities are inside the InT hub. Even though the DFD shows the squares, by default structure, as external entities, the green squares are external to the UAM database, but internal to the hub itself. In addition, it is not normal practice to make data flow lines with double arrowheads, but these, in this case, present a clear representation of the macro level 0 diagram. Separate uni-directional arrows exist in the level 1 DFD.



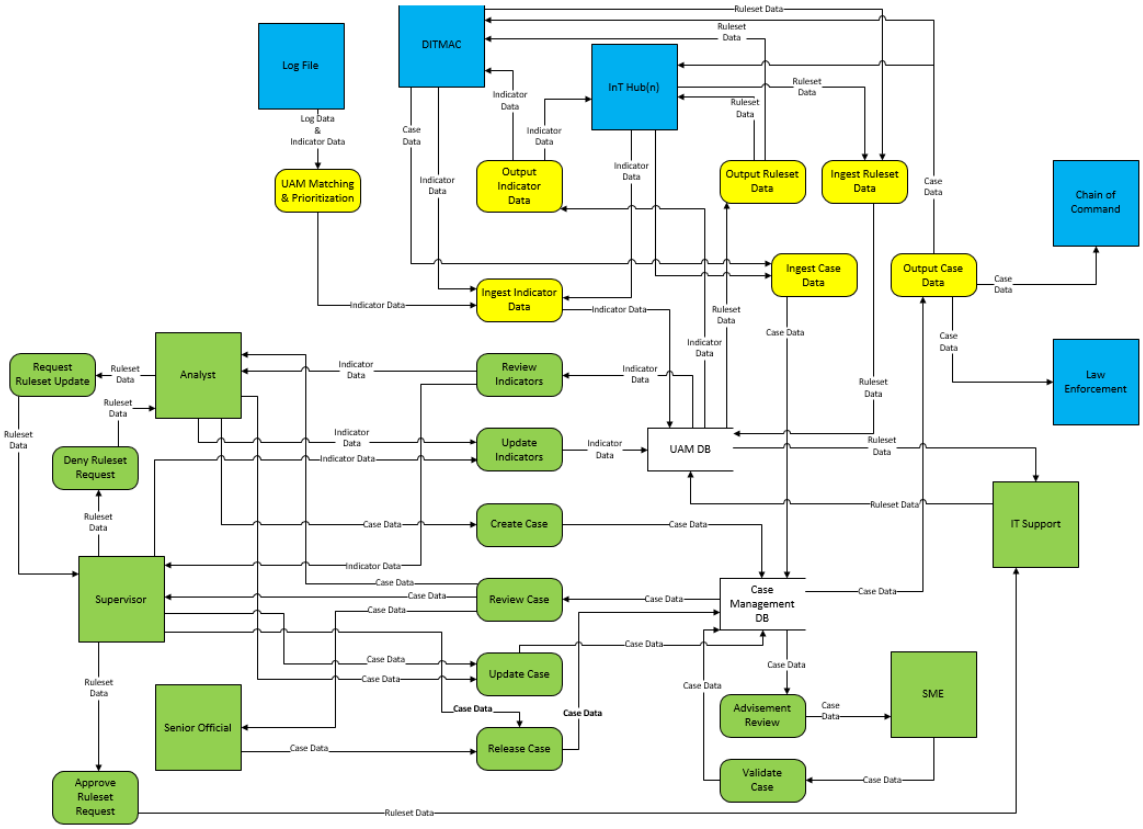
This DFD shows the entities internal to the hub (green), external to the hub (blue) and the InT hub operations as the center. The flow indicator arrows show the types of data passed.

Figure 18. Level 0 DFD for InT Hub Operations

This DFD shows the main elements mentioned earlier in this chapter. Although the diagram includes a UAM system, it does not represent a particular UAM vendor, but a generic UAM system that has industry-standard minimum capabilities. A Microsoft Visio supplemental, available from the Naval Postgraduate School upon request, represents the entire DFD of the InT hub, but this section presents an image version as well.

The work discovered during the workflow examination of the people, the technologies, and the processes was illuminating. It provided the knowledge needed to

create an accurate data flow diagram of how the data flows through the InT Hub, as shown in Figure 19. This data flow diagram captures the entities listed earlier in this chapter and the actual data types (taken from the artifacts of work tables within Section B of this chapter) that traverse from entity to entity.



This DFD shows the data flowing between the entities annotated within the level 0 DFD earlier in this chapter. The diagram captures the processes all entities perform as well as the data types moving throughout the system.

Figure 19. Level 1 DFD of InT Hub Operations

The level 1 DFD explores one level deeper than the level 0 DFD provided earlier in this chapter. The DFD includes all of the original entities along with their expected data types for a data sharing solution. In the yellow processes, a hub ingests and exports information from or to multiple hubs. Data translation challenges can occur within these processes. Since diverse data types at each hub is likely, there would have to be some translation capability at each hub for every hub connected to it. This creates an

exponentially difficult problem as the number of hubs increase. A solution would have to be able to model the data types and their relationships between each other in a semantic way in order to reduce the complexity of this problem.

The DFD captures all of the tasks covered within the workflow diagram earlier in this chapter. Every task element has a data input and output, which this diagram captures. The data inputs and outputs represent the artifacts of work generated by the InT hub system. The next section provides an examination of the different artifacts of work.

G. ARTIFACTS OF WORK

The artifacts of work presented by examining the people, technologies, processes, and their interactions led to some interesting insights. One of the most interesting elements was the small set of data types that surfaced. The level 1 DFD shows four data types within the system: indicator data, case data, ruleset data, and log data. The log data, from an inter-hub sharing perspective, is not as significant as the other three, since the UAM solution ingests the log data before it would share this data with other InT hubs. This leaves indicator data, case data, and ruleset data to consider in determining an inter-hub sharing solution.

The indicator data contains information about a single event recorded by a UAM client. This includes information like the time of the event, the user, and a description of the activity, among other information that is relevant to that single event. The data may have additional comments by an analyst or supervisor to expand upon the indicator. If a hub found this data and categorized it as suspicious, sharing that information with other hubs could allow those hubs to look for similar activity on their networks, if the indicator matched their environment model.

The case data contains one or more indicators and includes other information added by the analyst. This can include data collected from other sources, which helps to build the case for ultimate administrative or legal action. If InT hubs share this information, other hubs could see how the analyst analyzed and fused different data sources to create a credible picture of the suspicious activity. This would allow the other hubs to modify the analysis in their hub to be better poised to detect an insider threat.

The ruleset data contains UAM rules that, when matched to actual activities on the network, can begin the initial indicator of foul play by insiders. This information could be useful to other hubs because it could include elements such as tactics, techniques and procedures of insider threats and allow for shared knowledge of an insider threat's means to bypass security measures. If another hub has the same UAM tool, the language of the ruleset should be understandable. All a hub has to do is adjust the new rules so that they will be effective in the context of its own culture and environment. If the case management tools were different, an organization may have to implement a translation capability.

These three data elements make up the artifacts needed to link hubs together to allow sharing of useful data among the InT community. There currently is no readily shared information between InT hubs (M. Buckley, interview with author, March 17, 2016). Some of the concerns with sharing information, as learned from the NPS MMOWGLI game, are territorial and legal (Mascolo, 2016). The territorial challenge stems from organizations wanting to maintain control over their data. The legal challenge originates from security concerns over sharing information between hubs or outside of the InT hub architecture all together. Personally Identifiable Information (PII) is another concern, as the information requires that special protection measures be in place.

The challenge in creating a solution for sharing data and other relevant information among hubs is that it must be able to work with diverse data types and operate within the confines of the territorial and legal restrictions. A method must be identified which alleviates security concerns and allows organizations to maintain some semblance of control over their data. In addition, the solution must be able to operate within the confines of rules and regulations for the data types protected.

V. CONCLUSION AND FUTURE WORK

Insider threats are a significant threat to organizations and can be difficult to detect. Many performed studies, administered surveys, and examined datasets resulted in this conclusion. Insider threat organizations are still in their infant stages, learning as they go. With all of the advances in technology and after-thought system security solutions, an attacker has the advantage in the realm of cybersecurity.

The methods used in deterring, detecting, and mitigating insider threats include people, technologies, and processes, but unless these elements work in synchronous harmony, their success will be limited. In addition, combining knowledge of multiple organizations and sharing their successes and failures can lead to a better understanding of both the insider threat and how to minimize their effect. The InT community can only accomplish this by sharing information laterally with other community member organizations. There are challenges to this sharing of information, namely diverse data, regulations, and a desire to maintain control over one's own data. There exists a desire to share the information to better position InT hubs to be able to accomplish their mission, but the hubs must overcome these obstacles with any presented solution.

Cross-functional workflows and data flow diagrams provided information about the work and artifacts of work existing within an InT hub. The shareable elements of data that exist within a hub is a small set, but it is very powerful if shared properly. The solution must be able to share the data in a meaningful and lasting manner.

Future work based on the results presented here may include different semantic approaches to managing data among diverse data type sets. DITMAC is in an excellent hierarchical position to be a catalyst for positive InT community change within the DOD. Any solution that includes DITMAC and is able to gain support of the DSS leadership is likely to have the best chance of success.

The constant fight against the insider threat is not likely to cease; the adversary will always be there, waiting for their opportunity to strike. A strong insider threat deterrence, detection, and mitigation program will be the defense that keeps them at bay.

Together, the InT community stands the best chance to minimize the insider threat, but divided in silos of operation, the defenders do not stand a chance. The InT community must share this data and the tactics, techniques, and procedures, so that a daily dominance over the insider threat can be achievable.

APPENDIX. DSS RESPONSE TO INTERVIEW QUESTIONS

This appendix includes the interview responses of Mr. Randy Laylo, the Program Manager of the Defense Security Service Insider Threat Identification and Mitigation Center. Questions are in normal font and his responses to the questions are in boldface.

A. SYSTEMS

1. What SIEM solution does your organization use?

   a. [Classified]

2. What IDS/IPS does your organization use?

   a. CND shares policy violations

3. What UAM does your organization use?

   a. [Classified]

4. What other inputs serve as sources to the hub?

   a. Human resources

   b. Email pool

   c. Phone hotline

   d. Human capital management

   e. Lexis Nexis (Federal, not state)

5. What are the rates at which each of these sources provide alerts?

   a. 700–1000 data points / week

6. Do these alerts arrive in the analysis hub with assigned priorities?

   a. Alerts are color coded if they meet pre-programmed thresholds (“Special”)

   b. Low, Medium, High, Critical (based on physical and/or logical access)

7. What are the priority levels?

   a. Color coded (user configured)

8. What is the distribution of priority levels from each of these sources?

   a. Medium to Low are the majority (~30%)

   b. Very few High or Critical

   c. Most other alerts are not color coded (collection)

B. DATA

1. What categories of thresholds are important to the organization and why?

   a. Anonymized browser usage, Office of the Chief Information Officer triggers, financial indicators

2. How is the alert data presented to the organizational analyst?

   a. User interface [specifics are classified]

3. How does the organization expect an analyst to fuse information from additional sources?

   a. Analyst reaches out to SME support (through supervisor) and also accesses databases listed in input question (#4 above)

4. What means does the organization employ to allow for individual reports to be made as an input to the hub?

   a. Hotline, email

5. Does the organization handle alerts differently based on how they are received?

   a. High priority first, group email box is first come, first served

C. AUTOMATION AVENUES

1. What are the manual actions the organizational analysts do after initial input and before the case is created?

   a. Examine indicator for validity (remove false positives)

   b. Fuse data from other sources

   c. Create case for supervisor

2. What are the organizational policies on time spent on individual alerts?

   a. Text string matches = less than a minute per alert (~85% of the alerts)

b. Video capture = ~5 minutes (~3% of the alerts)

c. Classified word scrub = 2–3 minutes

d. Email attachments = 2–3 minutes

e. Email sent to foreign domain = 2–3 minutes

D. INFORMATION SHARING AND PROCESS

1. Does the organization have policies that limit sharing of certain information at certain stages of case management?

   a. Free sharing among the InT pillars (all SMEs sign a tailored Non-Disclosure Agreement)

2. What are the policies limiting this sharing?

   a. None to speak of

3. What are other organizational limitations on the processing of information?

   a. None to speak of

4. Does the organization alter data when it is sent out for consult / further review and in what way?

   a. No, full information disclosure (tailored Non-Disclosure Agreement)

E. ANALYST / WORKFLOW

1. How many analysts does the organization employ for how many clients?

   a. Two government analysts, one INFOSEC analyst, one contract analyst, one tool technical (DB management and ruleset creator) make up the primary hub. Tangential personnel support across DSS directorates.

2. Has the organization identified that the analyst workload is overwhelming the analysts?

   a. No

3. What is the organizational policy on establishing priorities of the data analyzed?

   a. Two primary categories:

i. Collection (all data that flags)

ii. "Special" (all data that triggers on pre-programmed high-level thresholds

4. How does the organization identify if a case is successful?

   a. Get to the left of the problem. It is not about walking someone out the door; it is about noticing an issue and notifying leadership to discuss with the person.

5. How much data comes into the organization analytic hub and how much is processed?

   a. 700–1000 points of data per week

6. What is the procedure for distributing the alerts to hub analysts?

   a. First come, first served. Analyst can make comments on alerts to be seen by all analysts. Color-coding is seen the same among analysts.

7. Are high priority alerts addressed before lower priority alerts?

   a. Yes

8. What are the possible outcomes for an analyst processing an alert (e.g., false alert, SME input required, forwarded for action, etc.)?

   a. False positive

   b. Case creation (definite violation)

   c. Enhanced monitoring (suspicious activity warranting more monitoring)

   d. SME involvement (advice, extra resource access)

9. What is the mean and Standard Deviation analyst processing time for each of these outcomes?

   a. Truly depends on the SME and the case. Some feedback is immediate; some can take months.

10. Does the analyst modify alert priority levels based on his analysis?

    a. Yes, analyst can raise or lower priority and then that is the attached priority until changed by another analyst or supervisor.

11. What is the organization's procedure on requesting Subject Matter Expert (SME) input?

    a. All analysts submit request for SME assistance to supervisor. Supervisor reaches out to SME. Response normally comes back to the original analyst if possible, but back to the supervisor if that analyst is unavailable.

12. Does the organizational process allow for direct liaison with the SME?

    a. Supervisor to SME, analyst never reaches out directly to the SME.

13. What is the mean and Standard Deviation processing time for each of the SMEs?

    a. Truly depends on the SME and the case. Some feedback is immediate; some can take months.

14. When the SME analysis is complete, is it returned to the analyst who initiated the query?

    a. When possible, but not guaranteed

15. When the analyst has completed his analysis, what is the organizational policy on where the analysis goes and what actions are taken on it?

    a. Based on the severity and urgency of the case, the speed of elevation may differ, but it goes from analyst to supervisor (if initial analyst is contractor, it goes first to the government analyst (more access) and then to the supervisor), to SME (if needed, returned to analyst or supervisor (if needed), elevated to the Chief of Staff (carbon copying Program Manager) and finally to Law Enforcement (811 referral) or other agency external to the housing agency. If the problem can be taken care of in-house (preferred), it is sent from InT hub supervisor to the supervisor of the person for action (disciplinary or remediation)

F. CUSTOMER NEEDS

Does the organization have specific needs for data if it were to be ingested from another hub to be used for efficient and effective analysis?

a. 100% financial disclosure as this shows the most effective personal stressors on people

b. Polygraph of 100% of the personnel

c. Regular expression that alert triggered on as well as the exclusion statements (e.g., A and B, but not C)

G. IT PERSONNEL QUESTIONS

1. How is the organization's InT source data stored?

   a. The InT UAM is housed in a database and the overall source data will be managed / housed in an "Incident Tracking" tool set that will allow the members of the hub the ability to access those items relevant to their position and subject area within the Hub and InT.

2. [How] is the data encrypted?

   a. Depending on where the data resides, AES is the standard, there are items in pace for Data at Rest in addition to the unique nature of how UAM is deployed in the enterprise. Beyond the encryption aspect, there are Role Based Access Controls in place as well.

3. How is the data backed up?

   a. Data will be backed up utilizing a combination of "Full" and "Incremental" backups of the system data.

SUPPLEMENTAL

There are two supplemental files included with this thesis: an InT Hub workflow diagram and an InT hub data flow diagram (DFD). These two files provide full versions of the portions included within the main body of the thesis. The supplemental files are available at the Naval Postgraduate School (NPS) in Monterey, California through the Dudley Knox Library upon request.

A. INT HUB WORKFLOW

One of the two supplemental files is a cross-functional workflow diagram that identifies the work that an InT hub must accomplish in order to complete a mission of insider threat mitigation. The workflow separates the work by role (horizontal rows) and by phase (vertical columns). The diagram shows the flow of work through the InT hub as well as who is performing the work and in which phase. The outcome of this diagram was to identify the tasks an InT hub must perform, which fed the next supplemental file, the InT hub data flow diagram (DFD).

B. INT HUB DATA FLOW DIAGRAM

The other supplemental file is a data flow diagram (DFD) that outlines the tasks and data types that originate from the cross-functional workflow diagram. The diagram represents the tasks as processes and the data flow arrows indicate the type of data that moves through the hub as well as the direction. The outcome of this diagram was an identification of the data types that the research needed in order to establish what requirements a data sharing solution would have to meet.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Ambler, S. W. (2000, July). Mapping objects to relational databases: What you need to know and why. Retrieved from http://www.ibm.com/developerworks/library/ws-mapping-to-rdb/ws-mapping-to-rdb-pdf.pdf

Balakrishnan, B. (2015, October 6). Insider threat mitigation guidance. Retrieved from https://www.sans.org/reading-room/whitepapers/monitoring/insider-threat-mitigation-guidance-36307

Brackney, R. C., & Anderson, R. H. (2004, March). *Understanding the insider threat: Proceedings of a March 2004 workshop.* Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/conf_proceedings/CF196.html

Cappelli, D., Moore, A., & Trzeciak, R. (2012). *The CERT guide to insider threats: How to prevent, detect, and respond to information technology crimes (theft, sabotage, fraud).* Westford, MA: Pearson Education, Inc.

Claycomb, W. R., Huth, C. L., Flynn, L., McIntire, D. M., & Lewellen, T. B. (2012). Chronological examination of insider threat sabotage: Preliminary Observations. *Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications, 3*(4): 4-20. Retrieved from http://isyou.info/jowua/papers/jowua-v3n4-1.pdf

Codd, E. F. (1970, June). A relational model of data for large shared data banks. Retrieved from http://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf

Defense Security Service. (n.d.). *DOD Insider Threat Management Analysis Center*. Retrieved July 21, 2016, from http://www.dss.mil/about_dss/ditmac.html

Department of Defense. (2014, Sep. 30). *The DOD insider threat program* (DODDirective 5205.16). Washington, DC: Author.

Elmasri, R., & Navathe, S. B. (2000). *Fundamentals of database systems* (3rd ed.). Reading, PA: Addison-Wesley.

Esposito, R., & Cole, M. (2013, August 26). How Snowden did it. Retrieved from http://www.nbcnews.com/news/other/how-snowden-did-it-f8C11003160

Exec. Order No. 13587, 3 C.F.R. 63811 (2011).

Finn, P., & Horwitz, S. (2013, June 21). U.S. charges Snowden with espionage. *The Washington Post.* Retrieved from https://www.washingtonpost.com/world/national-security/us-charges-snowden-with-espionage/2013/06/21/507497d8-dab1-11e2-a016-92547bf094cc_story.html

Gray, Jim. (1981). The transaction concept: Virtues and limitations. *Proceedings of the 7th International Conference on Very Large Databases*, *7*: 144–154.

Intel Security. (2015). Grand theft data: Data exfiltration study: Actors, tactics, and detection. Retrieved from http://www.mcafee.com/us/resources/reports/rp-data-exfiltration.pdf

Intel Security. (2016). Retrieved from: http://www.intelsecurity.com

Intelligence and National Security Alliance. (2013, September). A preliminary examination of insider threat programs in the U.S. private sector. Retrieved from http://www.insaonline.org/i/d/a/b/InsiderThreat_embed.aspx

Lee, R. M. (2015, August). The sliding scale of cyber security. Retrieved from https://www.sans.org/reading-room/whitepapers/analyst/sliding-scale-cyber-security-36240

Mascolo, R. (2016). *Insider threat operations: Is collaboration a key factor?* Master's thesis, Naval Postgraduate School. Monterey, CA.

Masters, G. (2015, November 19). Report: Insider threat more dangerous than external risks. Retrieved from http://www.scmagazine.com/report-insider-threat-more-dangerous-than-external-risks/article/455117/

Nakashima, E. (2011, May 4). Bradley Manning is at the center of the wikileaks controversy. But who is he? *Washington Post.* Retrieved from https://www.washingtonpost.com/lifestyle/magazine/who-is-wikileaks-suspect-bradley-manning/2011/04/16/AFMwBmrF_print.html

Obama, B. (2012, November 21). *National Insider Threat Policy* [Memorandum]. Washington, DC: The White House.

Ponemon Institute LLD. (2015). 2015 Cost of Cyber Crime Study: Global. Retrieved from https://ssl.www8.hp.com/us/en/ssl/leadgen/document_download.html?objid=4AA5-5207ENW&siebelid=12503&sectionid=pdf&returnurl=/us/en/secure/pdf/4aa5-5207enw.pdf&simpletitle=ponemon%20cyber%20report&subbu=tsg.software&parentPageName=3.0&analytics_page_name=3

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California