Carnegie Mellon University
Software Engineering Institute

NAVIGATING THE INSIDER THREAT TOOL LANDSCAPE: LOW COST TECHNICAL SOLUTIONS TO JUMP-START AN INSIDER THREAT PROGRAM

Derrick Spooner (Software Engineering Institute, Carnegie Mellon University)
George Silowash (Norwich University)
Daniel Costa (Software Engineering Institute, Carnegie Mellon University)
Michael Albrethsen (Software Engineering Institute, Carnegie Mellon University)

June 2018

Abstract

This paper explores low cost technical solutions that can help organizations prevent, detect, and respond to insider incidents, and provides the following:

- an overview of features and functionality associated with insider risk mitigation tools
- a taxonomy for high-level categories of insider threat tools
- a discussion of the relationship between the types of tools, the nuances of insider threat control deployment, and considerations for selecting, implementing, and operating insider threat tools

Introduction

Insider threat mitigation efforts involve the collection and analysis of a broad range of data. An effective insider threat program (InTP) incorporates a number of technical controls to assist with preventing, detecting, and responding to concerning behaviors and activity. These controls are often capabilities of tools that fall into one of five categories:

- user activity monitoring (UAM)
- data loss prevention (DLP)
- security information and event management (SIEM)
- analytics
- digital forensics

Commercial tools are available that address all of these categories. However, they are typically geared toward large enterprises, with purchase prices and implementation costs that are out of reach for many smaller organizations. This issue creates a barrier and a deterrent for many organizations that need to implement an InTP.

This report is intended for organizations that already have an established network security posture and would like to increase their InTP security posture with minimal software investment. It explores low cost tools available to organizations to help them jump-start the technical aspect of the InTP. These tools will still require physical hardware and expertise to install, manage, and maintain. Some of them might come with little to no technical support. Often, the developers of open source software contribute to a project on their own free time with little to no compensation. Some products do offer paid support, while others rely on a community of members to provide support to the project. Documentation might also be limited or non-existent in some cases. It is important to keep these factors in mind when considering an open source project. There are many popular evaluation methods, frameworks, and approaches to considering open source software [1]. Commercial vendors might also offer free or lower cost versions that lack certain features, only support a small number of users, or have other limited capacities. Exploring and using open source tools might be necessary in the early stages of program development before migrating or upgrading to commercially supported software.

Getting Your Program Started

Many organizations have already invested time and capital developing their IT infrastructures. Existing devices and applications should be evaluated to determine if they might help an organization enhance its InTP. For example, network firewalls often have additional functionality that can be activated either by purchasing the feature or simply enabling an existing capability, such as content filtering. This option can save an organization money and time.

Organizations should also review their current network topology to determine if it can effectively support an insider threat program. There might be opportunities to reposition devices in the organization to further enhance security with minimal or no compromise in functionality. Existing devices might support multiple network segments or virtual local area networks (VLANs) thereby allowing the organization to increase the security of InTP assets with little to no additional costs. This solution will also have the added benefit of increasing the return on investment of a particular device.

Overall, this report presents the tool categories that might already exist elsewhere in an organization outside of the InTP. An organization can use these high-level features and exemplars as a starting point for an in-depth software acquisition evaluation based on its current and future needs.

Table 1: Example Tool to Observable Mapping

Type	Observable	UAM	DLP	SIEM	Analysis	Forensics
Fraud	Network or Host Data Exfiltration	X	X	X	X	X
	Recruitment of insider via chat or email	X		X	X	
	Creation or use of fraudulent assets	X		X	X	
	Anonymous reporting				X	
	Social Engineering	X			X	
	Internal and/or external collusion	X	X			
Sabotage	Coworker conflict	X			X	

Type	Observable	UAM	DLP	SIEM	Analysis	Forensics
	History of rule violation				X	
	Disgruntlement or unmet expectations (demotion or termination)	X			X	
	Creation of unknown access paths (backdoor accounts)	X		X	X	X
	Deletion of logs	X		X	X	X
	Introduction of unauthorized code of software	X		X		X
Theft	Network or Host Data Exfiltration	X	X	X	X	X
	Physical Data Exfiltration (Print/Scan/Copy/Fax)	X		X		
	Announcement of resignation or termination	X			X	
	Access outside of need-to-know	X	X	X	X	
	Solicitation from external parties	X	X			
	Suspicious travel				X	

Tool Categories

These five categories are not comprehensive, but address observables for the three primary types of insider threat (Theft, Sabotage, and Fraud) as defined by Cappelli et al. [2]. Table 1 displays a mapping between common observables for each type of crime and the tool category capable of detecting that observable. When evaluating UAM, DLP, SIEM, and Analysis tools, it is worth examining them at the feature level, as there is a great deal of overlap between them. For example, both UAM and DLP software are generally capable of detecting content creation based on certain keywords or patterns, so it might not be necessary to deploy both.

User Activity Monitoring

The tool category for user activity monitoring is very broad and encompasses a variety of tools. The National Insider Threat Task Force states that user activity monitoring is “the technical capability to observe and record the actions and activities of an individual, at any time, on any device accessing U.S. Government information in order to detect insider threats and to support authorized investigations” [3, p. 2]. This particular requirement is not specific to U.S. government organizations.

Organizations should have visibility into host-based activities on their assets. This will not only help the organization prevent and detect malicious insiders, but it will also play a key role when an organization is responding to and investigating an incident.

Data Loss Prevention

Data loss prevention (DLP) tools must be able to identify, monitor, and protect data at rest, data in motion, and data in use. The tools need to employ deep content analysis and must be configurable to meet an organization’s unique business objectives and information security needs [4].

A DLP tool allows organizations to control how users interact with data. This feature might include policies that prohibit users from copying content to removable media or emailing it out of the organization. The DLP solution should also be capable of generating audit logs to help support incident investigation.

Security Information and Event Management

Security information and event management systems aggregate logs into a centralized repository and can perform automated analysis on those logs to discover trends and detect anomalies. According to the *Computer Security Handbook*, “A security incident and event management (SIEM) system provides an additional method for collection, aggregation, and consolidation of logs from many types of devices. The SIEM leverages baselining and configurable rules to correlate the logs and provide real-time incident-based alerting” [5, p. 26.23].

SIEM systems can help detect anomalies, which can lead to discovering potentially malicious insiders. System baselining and correlation perform a first order of rudimentary analysis that presents a more organized view of the raw log data. SIEM systems also aide in investigations by providing evidence that can be used for both internal incident response and external legal actions. Logs from critical devices, especially those that support the InTP, should be sent to the SIEM for centralized storage and analysis.

Analytics

Analytics tools extend the query and alerting functionality of the SIEM. They can implement advanced machine-learning and statistical techniques to uncover and alert on anomalous activity based on the following:

- threshold/volume-based anomalies
- user/role-based activity baselining
- previously unidentified patterns and trends

They can also provide additional advanced visualization capabilities such as charts and graphs that can make anomalies more visually apparent.

Digital Forensics and Investigations

“A Road Map for Digital Forensic Research” defines digital forensic science as “The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations” [6, p. 16]. Organizations should have digital forensic tools to support investigations and allow a properly trained individual to preserve, collect, and analyze digital artifacts on a system or device. These tools can be used to assist in the investigation of malicious insider actions and provide the necessary evidence for potential legal actions.

Notes on Tools

Implementation Costs

Organizations can benefit from using low cost or free tools. However, there are other costs associated with using any type of tool. Some low cost solutions might offer little to no support from the developers, requiring individuals to support the tool and understand how to use and troubleshoot it. The tools might require the purchase of additional hardware or other software in order for them to work effectively. In some cases, multiple tools might be needed to satisfy a particular requirement. One software application might do something particularly well, but another might be needed to fill in gaps.

Testing

Applications discussed in this paper must be tested in a non-production environment before they are deployed to a production system. This will allow the organization to assess the application and determine if it fits its needs. It will also allow the individuals charged with using and maintaining the product to become familiar with it, making the initial configuration of the software easier and less likely to cause issues in the production environment once deployed. These applications can also be tested in isolated virtualized environments that mimic the production environment. For example, Spooner et al. [7] present a testing environment implementation that could be used to explore these tools.

Risk Analysis and Cost-Benefit

There are tradeoffs to using low cost or open source software versus commercial software. Commercial software is typically supported by the company that developed it. It has likely gone through additional testing and other quality assurance procedures. Open source software may not have all of these benefits. Both commercial and open source solutions can still require the purchase of additional hardware and software to make the product work correctly and efficiently. It is important for organizations to evaluate their risk and conduct a cost-benefit analysis before implementing any commercial or open source solution.

Legal Issues

Legal counsel must be consulted before deploying technologies that could affect the privacy and legal rights of an employee. Any solution that monitors employee behavior, such as content filtering and email monitoring, must be evaluated by legal counsel before being implemented. This evaluation will allow the organization to identify any legal exposures these technologies create and implement appropriate mitigating controls to reduce the risk. Additionally, all the software licensing agreements should be reviewed before deploying or testing a solution. Some licensing agreements for free and closed source software prevent use in certain types of environments, such as using a tool for commercial purposes. In general, it is always a good idea to involve the legal team before implementing any new initiative.

Disclaimer

The tools listed in this document are examples of products that can be used to jump-start an insider threat program. The list is not exhaustive. CERT does not endorse or recommend these products specifically nor determine their suitability for use in any environment. Unless otherwise noted in this document, the software packages were not tested in a lab and information collected about a product for this report was derived from publicly available information.

User Activity Monitoring

User activity monitoring involves a broad range of tools. According to The National Insider Threat Task Force, the Committee on National Security Systems (CNSS) Directive Number 504 states that UAM is "the technical capability to observe and record the actions and activities of an individual, at any time, on any device accessing U.S. Government information in order to detect insider threats and to support authorized investigations." The same directive further states, "Each department/agency must have the following minimum capabilities to collect user activity data: key stroke monitoring and full application content (e.g., email, chat, data import, data export), obtain screen captures, and perform file shadowing for all lawful purposes. UAM data must be attributable to a specific user. The department/agency should incorporate this data into an analysis system capable of identifying anomalous behavior…" [3, p. 2].

Non-government organizations might wish to adapt the above definition and requirement after consulting with legal counsel and establishing the organization's tolerance for risk. UAM tool capabilities range from capturing the content of email messages and chats, to full screen and keyboard capture. These tools can focus on one particular monitoring capability or offer a suite of monitoring capabilities. There are several monitoring strategies described below that could be implemented to allow for the collection of data in support of the minimum capabilities.

- Client level: An application is installed on the client side to collect data. The client application can report the data back to a centralized collector for reporting and analysis. Client side applications can allow for the collection of data that could not otherwise be collected using the operating system's or monitored application's built-in settings or tools. However, client side monitoring can be susceptible to circumvention and detection by the end user.
- Server level: Software or settings enabled on a server that allow the collection of data in support of the minimum capabilities. Typically, software installed on a client interacts with a server to function. Examples include enterprise chat and email. Server-side software can use the hosted application's built-in tools or have supplemental applications installed to capture user activity.
- Network level: Infrastructure devices such as proxies, firewalls, and content filtering systems can collect data to support the monitoring of user activities. Other devices might need to be installed

at key ingress/egress points to collect additional information that installed infrastructure devices cannot.

Table 2: Minimum UAM Capabilities and Monitoring Points

Capability	Client	Server	Network
Key Strokes	X		
Content of chat	X	X	X*
Content of files and documents	X	X	X*
Screen capture of display	X		
Video capture of display activities	X		
Capture of file versions as they are edited	X	X	
Web browser activity	X		X
Clipboard (copy, cut, and paste) activity	X		
Files accessed	X	X	
Kernel processes	X		
Applications executed by user	X		
USB port activity	X		
Removable media activity	X		
Email content	X	X	X

Table 2 illustrates the minimum UAM capabilities and the monitoring strategies that can support the collection of this data. It should be noted that capabilities with the “X*” designation will be less effective due to encryption of any data that is in motion across the network.

Open Source HIDS SECurity (OSSEC)

Open Source HIDS SECurity (OSSEC) is a host-based intrusion detection system (HIDS). It is capable of monitoring changes to systems, such as changes to critical files or operating system configurations. It can be used on a variety of operating systems, including Microsoft Windows and Linux [8].

OSSEC can be used as a standalone package but is more easily managed in client-server deployment models. The software can be configured to monitor many aspects of a computer system. OSSEC is capable of monitoring processes and files.

Event logs and other files can be monitored for specific events or changes. Changes to the Microsoft Windows Registry can also be monitored. The Windows Registry contains a wealth of information that can be beneficial to the InTP. For example, OSSEC can be used to monitor when a new USB device is introduced to the system [9]. Note that if a USB device was previously connected to the computer being monitored, it might not be detected because registry values might not be created or updated. It is possible to remove the sub-keys of the USBSTOR registry hive by using a tool from Nirsoft known as “USBDeview”; however, before doing so, it should be tested on a non-production system [10].

Security Onion

Security Onion is a collection of tools combined into one Linux distribution that makes implementing monitoring of network traffic, through the use of intrusion detection systems and log collection, more manageable for organizations. Security Onion can be deployed quickly using its built-in setup wizard. Security Onion is also scalable; multiple sensors can be deployed and centrally managed. More information about Security Onion can be found at the following location:

https://security-onion-solutions.github.io/security-onion/

Squid Proxy Server and E2guardian

Squid proxy server is a software package that can be used to optimize the delivery of web content, which can help reduce bandwidth and provide additional features, such as logging [11]. E2guardian works in conjunction with Squid proxy to provide content filtering. It can filter not only by IP address or URL, but also by phrase matching. Both of these tools work together to produce logs that allow you to monitor user activity within an organization. Squid proxy can be found at the following location:

http://www.squid-cache.org/

E2guardian can be found at the following location:

http://e2guardian.org/

Intrusion Detection Systems

There are a number of intrusion detection systems (IDS) that can be used to monitor for various types of traffic or communications at key network perimeter ingress and egress points. For example, an IDS can be placed at the perimeter of a critical server enclave to monitor for specific types of attacks. These devices can be configured using rules developed by the organization's information security team, with the InTP providing guidance about the types of scenarios they should consider monitoring.

Some examples of IDS software that are freely available include

- BroIDS, available at https://www.bro.org/
- Security Onion, available at https://security-onion-solutions.github.io/security-onion/
- Snort, available at https://www.snort.org/
- Suricata IDS, available at http://suricata-ids.org/

Packet Capture

Organizations might need to capture and log all network activity to help assist in incident response. However, full packet capture comes with a cost. An organization needs to determine where to position the device(s) and how much data they need to keep. The amount of data flow that passes through the sensor and is captured will determine the storage requirements as well. Hardware that is capable of capturing and storing the packets at a sufficient speed will also be needed. Having network data available can help an organization determine what events, at a network level, contributed to a security incident. A side benefit of a packet capture device is the ability to help assist network engineers with troubleshooting. Many IDS devices offer the ability to capture packets when a rule is triggered. This can help when an incident is being investigated.

Some examples of packet capture tools are described below.

- Tcpdump is a Linux utility that can capture and analyze network traffic [12]. Tcpdump could be placed onto an appropriately sized server to monitor and collect network activity. More about tcpdump can be found at the following location:
    - http://www.tcpdump.org/
- NetworkMiner is a Microsoft Windows-based application that can be used to capture and analyze network traffic using a GUI to make analysis easier for the analyst. More information about NetworkMiner can be found at the following location:
    - http://www.netresec.com/?page=NetworkMiner
- Wireshark is a packet capture and analysis tool designed for both Microsoft Windows and Linux. It offers a visual interface to help visualize packet data [13]. More about Wireshark can be found at the following location:
    - https://www.wireshark.org/

Data Loss Prevention

Data loss prevention (DLP) technologies generally protect data from leaving systems through unauthorized channels. When considering DLP technologies for use in an environment, you must consider the three types of data the system will be monitoring and protecting: data at rest, data in motion, and data in use.

Data at rest refers to data in storage awaiting use. Typically, data at rest refers to data stored on hard disk drives (HDD), solid state drives (SSD), removable media, or backup media, such as tapes. One of the more common ways to protect data at rest is through the use of encryption. Organizations can designate the use of encryption for certain types of data. For example, personally identifiable information (PII) or protected health information (PHI) might require the use of encryption depending on where it is stored. If the data leaves one storage location and is later stored at another location, such as data leaving a server for storage on removable media, the DLP solution might mandate the use of encryption on the removable media. The DLP solution can also check to verify that the media being used to store the data is authorized. For example, DLP solutions can enforce the use of particular USB flash drives with certain serial numbers or from specific manufacturers.

Data in motion is data flowing on the organization’s networks. DLP systems for data in motion can include combinations of hardware and software sensors at critical enclave ingress and egress points. For example, sensors can be deployed in front of servers that contain critical data assets, or between two network segments with different levels of trust. As data leaves or enters the enclave, it is checked to ensure it complies with the organization’s information security policies.

Data in use can be thought of as data that is being manipulated by a system or end user. It can also include the creation, modification, or deletion of data on an endpoint, such as a workstation or mobile device. The most common type of DLP system the Insider Threat Center at CERT sees involves systems deployed to prevent data exfiltration through the use of removable media, such as USB flash drives and other similar devices.

The “Common Sense Guide to Mitigating Insider Threats” states that “Organizations must understand not only their physical assets, but also their information assets and where they keep their most valuable and sensitive information and equipment. Physical assets, such as servers and workstations, are more easily tracked and protected. Data may be more difficult to track, but to protect it, organizations must understand the types of data they process, where they process it, and where they store it” [14, p. 31]. This level of security can be difficult for organizations to manage. The Insider Threat Center has seen organizations in the financial sector that give employees tools to scan their local workstations for sensitive data so that they can properly secure it, either by moving it to approved storage or by securely deleting the data. It is important for organizations to know where their most sensitive data resides so that they can best protect it.

OpenDLP

OpenDLP is a client/server-based tool that can scan endpoints for sensitive data. The client portion of the application is a service that resides on user workstations that scans the workstation based on settings pushed to it from the OpenDLP server. The OpenDLP server manages the results of the scans. OpenDLP also has the ability to scan Microsoft SQL Server and MySQL databases for sensitive information. It should be noted that OpenDLP does not prevent data loss, but instead identifies where sensitive data lives in your organization. More information about OpenDLP can be found at the following location:

https://github.com/ezarko/opendlp

MyDLP

MyDLP is a data loss prevention solution that has both a free community edition and a paid enterprise edition. The community edition has a limited feature set, while the enterprise version includes additional features and commercial support. The MyDLP website, http://www.mydlp.com, does not provide information to compare the two versions; however, the Internet Archive contains a version of their website with information about how the two versions compare. This comparison might not be accurate today, but it does provide an idea of the features the software offers. An archived version of the website can be found at the Internet Archive at the following location:

https://web.archive.org/web/20131201103303/http://www.mydlp.com/features/

While there are few open source tools in the DLP tool space, the tools available can be leveraged to help an organization identify and protect their sensitive information while initially establishing their insider threat program.

Security Information and Event Management (SIEM)

Security information and event management systems can be one of the most important components of an insider threat program. These systems receive log information from various devices across the enterprise. Many products aimed primarily at log collection can be configured to have SIEM-like functionality, or they can offer it via add-on products or licenses. Absent a full-fledged solution, alerting on configured log queries can fulfill some requirements of a SIEM. Insider threat programs typically have access to the enterprise SIEM. However, since the InTP mission differs from the typical network operations and security mission, the InTP typically creates and uses its own specific rule sets.

Figure 1 illustrates the vast amount of data that an insider threat program should analyze. A SIEM can help insider threat programs by consolidating logs into a central location and automatically prioritizing events, making those with a higher priority more visible to an analyst for action.

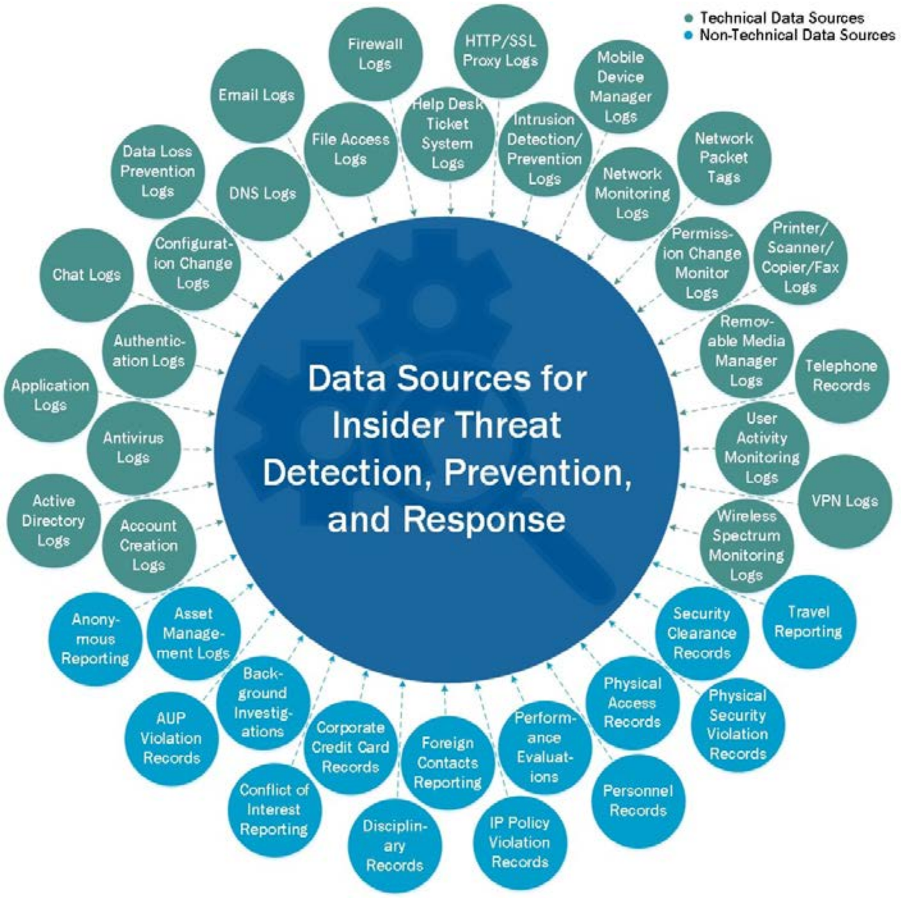
● Technical Data Sources
● Non-Technical Data Sources
Data Sources for Insider Threat Detection, Prevention, and Response
Firewall Logs
HTTP/SSL Proxy Logs
Mobile Device Manager Logs
Email Logs
File Access Logs
Help Desk Ticket System Logs
Intrusion Detection/ Prevention Logs
Network Packet Tags
Data Loss Prevention Logs
DNS Logs
Network Monitoring Logs
Printer/ Scanner/ Copier/Fax Logs
Chat Logs
Configurat-ion Change Logs
Permiss-ion Change Monitor Logs
Authentic-ation Logs
Remov-able Media Manager Logs
Telephone Records
Application Logs
Antivirus Logs
User Activity Monitoring Logs
VPN Logs
Active Directory Logs
Account Creation Logs
Wireless Spectrum Monitoring Logs
Travel Reporting
Anony-mous Reporting
Asset Manage-ment Logs
Security Clearance Records
Back-ground Investig-ations
Physical Access Records
Physical Security Violation Records
AUP Violation Records
Corporate Credit Card Records
Foreign Contacts Reporting
Perform-ance Evaluat-ions
Conflict of Interest Reporting
Personnel Records
Disciplin-ary Records
IP Policy Violation Records

Figure 1. Data Sources

A SIEM should have an easy-to-use interface and be highly configurable for both the organization and the analyst. AlienVault states that logs are valueless unless subjected to regular and random review, with follow-up if anomalies are detected. It is unrealistic to expect an individual to pore over voluminous log files on a daily basis. However, log aggregation and correlation technology can be employed to provide an additional layer of confidence as anomalous activity across systems can be related—potentially identifying an attack pattern or other irregular activity that would not be apparent from a single log [15].

It is not enough to simply install a SIEM and have logs sent to it. A typical SIEM can process hundreds to hundreds of thousands of events per second. With such a large volume of data, the SIEM rules must be finely tuned and the system configured appropriately to help determine which events are important to both the InTP and the organization's mission. The SIEM will process the events, categorizing and correlating them according to those specific rules. It can also be configured to email high priority alerts to insider threat program staff. InTP staff should have the ability, via dashboards or other means, to review and explore all events captured by the SIEM.

OSSIM

OSSIM by AlienVault provides event collection, normalization, and correlation [16]. AlienVault also offers a commercial version of this product with more features, such as allowing multiple users with role-based access control (RBAC), tiered architecture, and customizable reports. A comparison of the open source and commercial product can be found at the following location:

https://www.alienvault.com/products/compare-ossim-to-alienvault-usm

LOGalyze

LOGalyze is a SIEM that is free to download. While not open source, it is free to use in an enterprise environment. The software has the ability to ingest various log sources, such as files, SNMP, and other system logs. The tool can export reports to a variety of formats and can generate alerts when one or more events meets certain user-defined criteria. Additionally, the tool comes pre-configured with some compliance reports, including HIPAA, PCI-DSS, and Sarbanes Oxley. More about LOGalyze, including how to download and install it, can be found at the following location:

http://www.logalyze.com/

Enterprise Log Search and Archive (ELSA)

ELSA is a free and open source log aggregation and search tool. It can ingest logs from most common sources, normalize them, and provide the user with a searchable database. ELSA provides a custom query capability, and any query can be saved as an alert. While ELSA does not have many “point and click” reporting capabilities or dashboards, it is a highly customizable tool that can be configured with any number of particular searches or alerts. It is also worth noting that the aforementioned Security Onion tool suite also includes ELSA. More information, and the software itself, can be found at the following location:

https://github.com/mcholste/elsa

The Elastic Stack

Formerly referred to as Elasticsearch, Logstash, and Kibana (ELK), the Elastic Stack is another open source tool suite that provides log aggregation and reporting, and it also offers a subscription service for technical support. Much like ELSA, the Elastic Stack can ingest most common logging sources and provide a configurable platform for searches, alerts, and visualization. More information about the Elastic Stack, including how to download it, can be found at the following location:

https://www.elastic.co/

Analytics

Analytics allow the insider threat team to discreetly identify anomalies and analyze potential insider threat activity. The InTP collects a tremendous amount of data, often by a SIEM. Once data is in a repository, it must be analyzed in order to be of use to the InTP. This capability belongs with the InTP hub. The insider threat hub is a centralized capability for insider threat analysis and response. Some of the capabilities of the hub include

- collecting, correlating, and aggregating data from disparate sources
- developing, deploying, and refining indicators of potential insider activity
- evaluating detected instances of potential insider activity
- providing supporting information to incident investigators and responders

It should be noted that the InTP hub is not a specific tool; it is a collection of tools and capabilities that support the InTP. The hub helps to paint a picture of the "whole person." That is, no one indicator can identify a potential malicious insider. It takes multiple indicators to help establish whether or not a person should be of interest to the InTP.

Figure 1 depicts all of the data sources that should feed into the InTP hub. This data might be scattered across the enterprise. Ultimately, this data should be centrally collected and maintained. However, laws, regulations, and organizational policies dictate how this information is collected and managed. The InTP needs access to these data sources to help paint a more complete picture of an individual and understand all aspects of the inquiry.

To jump-start an InTP, the organization must identify the types of scenarios that it wants to defend against, starting small and growing the program once it has perfected a few capabilities. Once the organization has selected scenarios, the organization needs to determine what data feeds would allow them to prevent, detect, and respond to the identified scenario. The data feeds will then need to be examined to determine if they contain sufficient information to help support the identification of malicious insider activity. In some cases, the data feed might be insufficient and additional fine tuning might be needed. For example, certain account actions might be missing from a log, but can be made available easily by adjusting account audit policies. Once the organization is able to identify malicious insiders given the selected cases, the organization can then build upon this capability by selecting additional scenarios it wants to include in the InTP.

Analytic tools that are part of insider threat programs fall into two categories. Machine learning is the ability for computers to learn without explicit programming [17]. Predictive analytics, as described by Wayne Eckerson, is "a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events. Unlike other BI technologies, predictive analytics is forward-looking, using past events to anticipate the future" [18, p. 1].

In many cases, these tools are combined to monitor the vast amounts of data the InTP collects. The InTP hub requires the assistance of these tools to sift through the data to identify patterns.

The two tools described below can help the InTP hub analyze the data it receives. These tools might require expertise from people across the organization to be integrated.

Weka

Weka is a tool that can be used to analyze large data sets using machine learning [19]. Machine learning can be applied to large data sets to identify patterns of activity that may be of interest to an InTP. More information about Weka can be found at the following location:

http://www.cs.waikato.ac.nz/ml/weka/

RapidMiner

RapidMiner is a predictive analytics tool designed to help organizations predict events. The tool can be used to analyze events to help the end user make more informed decisions about events that can occur. This tool can be beneficial to InTPs because it allows them to identify people of interest before they become malicious insiders. More information about the tool can be found at the following location:

https://rapidminer.com/

Digital Forensics and Investigations

A key part of any successful insider threat program is the ability to conduct a sound forensic examination of digital evidence. Malicious insiders use technology to carry out their crimes, and evidence of these crimes can be found on the systems they use. However, preserving, recovering, analyzing, and reporting this evidence is the biggest challenge many organizations face. The costs of commercial tools for conducting a digital forensic investigation can be significant.

It is also important to note that an untrained individual can create a liability to an organization attempting to conduct an investigation. Valuable evidence can be lost forever, exculpatory evidence can be missed, and data can be misinterpreted. This creates a risk to the organization by exposing it to lawsuits and can undermine the legal process. Therefore, before an organization decides to implement a digital forensics capability, careful consideration must be given to the organization's risk tolerance, existing capabilities, and the digital forensics knowledge of current staff. An individual with deep technical knowledge, such as a systems or network administrator, might be good at their job, but that does not make them an investigator without the proper training and experience. Management should work with their organization's legal counsel when developing a digital forensics capability. Senior leadership must also understand the other costs associated with a digital investigation capability, such as

- specialized hardware (e.g., write blockers, storage, forensic workstations, servers)
- software licensing and annual maintenance renewals
- evidence storage facilities
- annual training for staff

It can be more beneficial and cost effective to outsource a digital investigation than to implement the capability at a particular organization.

Careful consideration must be used when deploying a tool to aide in a digital investigations. A tool should be tested in a controlled environment so that it can be repeatable and reproducible. The National Institute of Standards and Technology provides the following related definitions:

- repeatability: precision under repeatability conditions
- repeatability conditions: conditions where independent test results are obtained with the same method on identical test items in the same laboratory by the same operator using the same equipment within short intervals of time
- reproducibility: precision under reproducibility conditions
- reproducibility conditions: conditions where test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment

As applied to computer forensic testing, repeatability is defined as the ability to get the same test results on the same testing environment (e.g., the same computer, disk, mode of operation, and so forth). Reproducibility is defined as the ability to get the same test results on a different testing environment (e.g., a different PC, hard disk, operator, and so forth) [20].

It is important to note that courts do not approve digital forensics tools. This term is often confused with “court accepted.” *Computer Forensics* defines the Frye Standard as “Forensic tools, techniques, procedures, and evidence [that] are admissible in court if they have a ‘general acceptance’ in the scientific community” [21, p. 383]. Any tool, whether open source or closed source, must undergo testing to ensure it is able to produce accurate, reliable, repeatable, and reproducible results.

When introducing a tool or technique to their investigations, digital investigators should keep the Daubert standard in mind, which is defined as a “Standard used by a trial judge to make a preliminary assessment of whether an expert’s scientific testimony is based on reasoning or methodology that is scientifically valid and can properly be applied to the facts at issue. Under this standard, the factors that may be considered in determining whether the methodology is valid are: (1) whether the theory or technique in question can be and has been tested; (2) whether it has been subjected to peer re-view and publication; (3) its known or potential error rate; (4) the existence and maintenance of standards controlling its operation; and (5) whether it has attracted widespread acceptance within a relevant scientific community” [22].

Before using any forensic tool or technique, consult with legal counsel to ensure that it will withstand legal scrutiny and is acceptable to use.

FTK Imager

FTK Imager is a tool used by incident first responders to preserve evidence before it is destroyed. While it is not an open source tool, it is available free from AccessData. FTK Imager should not be confused with AccessData's Forensic Toolkit (FTK). FTK Imager is for acquiring evidence that will preserve the data in a manner that meets evidence admissibility requirements. FTK Imager is not for conducting in-depth forensic examinations.

FTK Imager allows you to preview data on a device and make a forensically sound image of the evidence without making changes to the device being imaged [23]. FTK Imager should be used in conjunction with a hardware write-blocking device to prevent inadvertent changes to evidence. The write-blocking device will prevent data from being written to the device connected to it for imaging and analysis. It is important to note that solid state drives (SSDs) present new challenges to the forensics community, such as garbage collection and wear leveling, which may cause hash values not to match. Further discussion of this technology is outside the scope of this paper, however. FTK Imager can be obtained from the following location:

http://accessdata.com/product-download?/support/product-downloads

Once a forensic image of a subject's computer or storage media is obtained, a duplicate copy of the image should be created. The duplicate image can then be examined using other forensic tools.

Autopsy

Autopsy is a GUI-based software application that allows an analyst to examine various types of evidence that might be involved in an incident. Additional plugins can be developed to enhance Autopsy's capabilities [24]. Autopsy has many of the same features as some of the commercial digital forensics packages, such as AccessData's Forensic Toolkit (FTK) and Guidance Software's EnCase Forensic. A new feature available in Autopsy is the ability for multiple people to collaborate on a case at one time.

Autopsy is a powerful tool that can be used by organizations to investigate incidents. It has a robust feature set that can be used to examine collected evidence to help determine the cause of an incident and possibly what the subject did. Autopsy can be downloaded from the following location:

http://www.sleuthkit.org/autopsy/

Volatility

Volatility is a framework for analyzing memory captures from a computer system. Memory captured from a computer system can be used to help further an investigation. For example, memory might contain information about the processes that were executing on a system. If data encryption is in use on the system, the encryption keys may be found in memory. This information can be used to help paint a picture of what a subject might be doing should he or she be suspected of malicious insider activity. More information about Volatility can be found at the following location:

http://www.volatilityfoundation.org/

SANS Investigative Forensic Tookit (SIFT)

The SIFT workstation is a compilation of free and open source forensics tools contained in a Linux virtual appliance. The toolkit is preconfigured and ready to use as a virtual appliance, or it can be manually built on top of an Ubuntu machine using scripts to compile and build the workstation.

The SIFT workstation contains a variety of tools that can be used to conduct an investigation or respond to an incident. A complete investigation can be conducted with the tools included in the workstation. The SIFT workstation can be downloaded from the following location:

http://digital-forensics.sans.org/community/downloads#locations

CERT Forensics Tools

CERT offers an extensive set of forensics tools to help investigators. The SEI states that “The CERT Linux Forensics Tools Repository provides many useful packages for cyber forensics acquisition and analysis practitioners” [25]. The tool repository instructions can be found by accessing https://forensics.cert.org/. Additionally, CERT provides a virtual machine based appliance, called Appliance for Digital Investigation and Analysis (ADIA), which enables an investigator to use open source tools to conduct an investigation. ADIA is a VMware-based appliance used for digital investigation and acquisition and is built entirely from public domain software. Among the tools contained in ADIA are Autopsy, the Sleuth Kit, the Digital Forensics Framework, log2timeline, Xplico, and Wireshark. Most of the system maintenance uses Webmin. The appliance runs on Linux, Windows, and Mac OS. Both i386 (32-bit) and x86_64 (64-bit) versions are available. ADIA is available to the public and is designed for small-to-medium sized digital investigations and acquisitions. It provides an alternative method for conducting digital investigations [26]. More information about ADIA and download instructions can be found at the following location:

http://www.cert.org/digital-intelligence/tools/adia.cfm

CERT also offers some standalone tools that an investigator can use to supplement their investigation, described below.

- AfterLife permits the collection of physical memory contents from a system after a warm or cold reboot.
- DINO is a lightweight front end for network visualization and uses the open source net-work monitoring tools SiLK and SNORT to create an easy-to-use dashboard for situational awareness.
- LATK is a collection of command line and web-based tools for use in incident response and long-term analysis of web server and proxy server log data [27].

More information about these tools and how to download them can be found at the following location:

http://www.cert.org/digital-intelligence/tools/

Other Open Source Linux Distributions

There are numerous Linux distributions that contain standalone forensic tools, and some distributions contain complete forensic toolkits. One example of a complete Linux distribution is PALADIN.

PALADIN is a modified “live” Linux distribution based on Ubuntu that simplifies various forensics tasks in a forensically sound manner via the PALADIN Toolbox. PALADIN is a complete solution for triage, imaging, examination and reporting [28]. More information is available at the following location:

https://www.sumuri.com/product/paladin-for-linux-2/#

Summary

Insider threat programs use five categories of tools to help their organizations prevent, detect, and respond to malicious insider activity. These categories include

- user activity monitoring (UAM)
- data loss prevention (DLP)
- security information and event management (SIEM)
- analytics
- digital forensics

Organizations should include tools from each one of these categories as a minimum requirement for building an insider threat program. These tools will allow organizations to better understand how employees interact with systems and data. User activity monitoring tools allow organizations to understand how employees interact with all endpoints in their environment. These tools monitor how an employee uses company-owned assets.

Data loss prevention tools protect the organization’s data by leveraging technology to enforce where and how data is stored and who may access it. Data loss prevention also enforces the organization’s data classification policies and can take action to prevent data from being stored or transmitted in an unapproved manner.

Security information and event management systems collect and manage logs from various devices across the enterprise and help identify events that might be of interest to the organization. These systems help organizations digest large volumes of information and provide alerts on events of interest to the InTP.

Analytic tools leverage the data in security information and event management systems to discover patterns and trends in data that might be useful for identifying malicious insider threat behavior. Analytic tools help organizations process large volumes of information and assist in determining if any actionable intelligence is contained in the data.

Digital forensics and investigation tools allow an organization to respond to a malicious insider incident. These tools help the organization determine if and how an incident occurred. Information gleaned from these tools can be used to help strengthen the organization's security posture by incorporating lessons learned from the incident.

Organizations looking to start an insider threat program might have limited budgets for getting the technical part of the program off the ground. Therefore, organizations should look to leverage technology they already have by considering how it can help an insider threat program. If existing tools are insufficient and the organization has a limited budget, low cost tools should be evaluated to determine if they will help fill gaps in the InTP. While the tools might be low cost, ongoing support and maintenance of the tools is another factor the organization must consider when deciding to implement a particular tool.

Future Work

Future research in the area of insider threat tool usage includes methods to further tune and evaluate alerts and rules for UAM, SIEM, DLP, and analytic tools, increasing the scope of data input to these tools, and general academic validation of the utility of machine learning algorithms. Specifically, research could be sensibly applied toward minimizing false positive rates in the alerting tool categories to combat analyst fatigue and improve the overall effectiveness of these tools.

While tools in the UAM, SIEM, DLP, and analytic categories generally have some level of out-of-the box rules, they can be overly reactive and require great deal of investment in tuning and creating those rules. Tagging and protecting data besides PII and PHI with a DLP system generally relies on a deep understanding of the distinct intellectual property generated across organizational components. Advancements for DLP software include studying the best approach for obtaining this understanding of company IP, or better mechanisms to delegate the task to end users without degrading the user experience or productivity.

With regard to SIEM and analytic data sources, the field could benefit from exploring the specific value and utility of including additional sources for analysis. Analysis tools in the market also have varying approaches for assigning aggregate risk models including rule-based detection, supervised, and unsupervised machine learning detection. The impact and validity of these approaches have not been researched in detail in operational settings. Finally, future research for forensic tools includes studying the effectiveness of operationalizing the tools and techniques in a more real-time detective manner in addition to their traditional response application.

References

[1] K.-J. Stol and M. A. Babar, "A Comparison Framework for Open Source Software Evaluation Methods," in Open Source Software: New Horizons, Notre Dame, 2010.

[2] D. M. Cappelli, A. P. Moore and R. F. Trzeciak, *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes*, 1st ed., Addison-Wesley Professional, 2012.

[3] National Insider Threat Task Force, "Clarification of Enterprise Audit Management (EAM), User Activity Monitoring (UAM), Continuous Monitoring, and Continuous Evaluation," 14 03 2014. [Online]. Available: http://www.ncsc.gov/nittf/docs/EAM_UAM_and_Continuous_Monitoring_Definitions-Signed.pdf. [Accessed 17 12 2015].

[4] B. Banner, "Security Analyst to DLP Vendors: Watch Your Language," CSO Online, 02 June 2009. [Online]. Available: http://www.csoonline.com/article/2124056/security-industry/security-analyst-to-dlp-vendors--watch-your-language.html. [Accessed 22 December 2015].

[5] J. Opatrny, "Gateway Security Devices," in *Computer Security Handbook*, 6th ed., Hoboken, NJ: John Wiley & Sons, Inc., 2014, p. 26.23.

[6] G. Palmer, "A Road Map for Digital Forensic Research," Digital Forensics Research Workshop (DFRWS), Utica, 2001.

[7] D. Spooner, R. Ditmore and D. Costa, "NeedleStack: Infrastructure for Insider Threat Control Development, Testing, and Training," Software Engineering Institute, 2018.

[8] OSSEC Project Team, "OSSEC: About," 2015. [Online]. Available: http://ossec.github.io/about.html. [Accessed 4 January 2016].

[9] OSSEC Project Team, "Detecting USB Storage Usage," 2015. [Online].

Available: http://ossec.github.io/docs/manual/monitoring/process-monitoring.html#detecting-usb-storage-usage. [Accessed 4 January 2016].

[10] N. Sofer, "USBDeview v2.51 - View all installed/connected USB devices on your system," 2015. [Online]. Available: http://www.nirsoft.net/utils/usb_devices_view.html. [Accessed 4 January 2016].

[11] squid-cache.org, "Squid: Optimising Web Delivery," [Online]. Available: http://www.squid-cache.org/. [Accessed 11 March 2016].

[12] TCPDUMP & LIBPCAP, "TCPDUMP & LIBPCAP," [Online]. Available: http://www.tcpdump.org/. [Accessed 11 March 2016].

[13] Wireshark, "About Wireshark," [Online]. Available: http://www.wireshark.org. [Accessed 11 March 2016].

[14] G. Silowash, D. Cappelli, A. Moore, R. Trzeciak, T. J. Shimeall and L. Flynn, "Common Sense Guide to Mitigating Insider Threats, 4th Edition," Software Engineering Institute, Pittsburgh, 2012.

[15] D. J. Johnson, N. Takacs and J. Hadley, "Securing Stored Data," in Computer Security Handbook, 5th ed., John Wiley & Sons, Inc., 2009.

[16] AlienVault, "AlienVault OSSIM: The World's Most Widely Used Open Source SIEM," [Online]. Available: https://www.alienvault.com/products/ossim. [Accessed 22 02 2016].

[17] P. Simon, *Too Big to Ignore: The Business Case for Big Data*, Hoboken: John Wiley & Sons, Inc., 2013.

[18] W. Eckerson, "Predictive Analytics: Extending the Value of Your Data Warehousing Investment," 2007.

[19] Machine Learning Group at the University of Waikato, "Weka 3: Data Mining Software in Java," [Online]. Available:

http://www.cs.waikato.ac.nz/ml/weka/. [Accessed 09 03 2016].

[20] National Institute of Standards and Technology, "General test methodology for computer forensic tools," United States Department of Commerce, Washington, DC, 2001.

[21] M.-H. Maras, *Computer Forensics*, Burlington: Jones & Bartlett Learning, 2015.

[22] Cornell University Law School, "Daubert Standard," [Online]. Available: https://www.law.cornell.edu/wex/daubert_standard. [Accessed 11 March 2016].

[23] AccessData Group, LLC, "AccessData FTK Imager User Guide," 21 March 2012. [Online]. Available: https://ad-pdf.s3.amazonaws.com/Imager%203_1_4_UG.pdf. [Accessed 28 February 2016].

[24] B. Carrier, "Open Source Digital Forensics," [Online]. Available: http://www.sleuthkit.org/index.php. [Accessed 28 February 2016].

[25] Software Engineering Institute, "Linux Forensics Tools Repository," [Online]. Available: https://forensics.cert.org/. [Accessed 3 March 2016].

[26] Software Engineering Institute, "Appliance for Digital Investigation and Analysis (ADIA)," [Online]. Available: http://www.cert.org/digital-intelligence/tools/adia.cfm. [Accessed 3 March 2016].

[27] Software Engineering Institute, "Digital Intelligence and Investigation Tools," [Online]. Available: http://www.cert.org/digital-intelligence/tools/. [Accessed 3 March 2016].

[28] Sumuri LLC., "PALADIN (64-bit) – Version 6.08," [Online]. Available: https://www.sumuri.com/product/paladin-for-linux-2/#. [Accessed 3 March 2016].

Contact Us

Software Engineering Institute
4500 Fifth Avenue, Pittsburgh, PA 15213-2612

Phone: 412/268.5800 | 888.201.4479
Web: www.sei.cmu.edu
Email: info@sei.cmu.edu

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

DM18-0814