Introduction

Due to its clear benefits in detection, threat hunting has garnered the attention of many organizations. The primary focus of threat hunting is detecting attacks missed by other security controls. Threat hunting also allows us to address higher levels of the Pyramid of Pain,1 making the adversary’s life a lot harder. As a bonus, most of the techniques used in threat hunting scale well even for large environments, making it a viable solution for organizations of all sizes.

There are many existing definitions for threat hunting and some of them are vague. SANS defines threat hunting as a process using new information on previously collected data to find signs of compromise evading detection. Usually, it is a very manual and human-centric activity. It takes a proactive approach to detection; thus it is not based on signatures. The output of threat hunting either feeds directly into the incident response process if something malicious is detected or provides input for security monitoring resulting in new detection methods.

Implementing and managing threat hunting in an organization can be a daunting task. In this paper, we focus on the essentials when it comes to hunting for threats by providing recommendations on techniques, organizational structures, required data sources and tools to create a successful threat hunting program.

---

Before diving into the details of a threat hunting program, let’s address some of the confusion and misconceptions typically associated with it.

“Threat hunting is something very new.”
Many mature organizations already implement some form of a threat hunting program, and the concept is originally from the military.

“It requires a lot of different data types and resources and needs special, expensive tools.”
Pulling a few log files from a data source such as a web proxy server and analyzing them using command line tools and scripts can be a particularly effective way of threat hunting. Of course, as an organization’s threat hunting program matures, centralized log collection and analysis could be introduced to the process.

“Threat hunting is something highly sophisticated and requires a lot of expertise.”
Implementing simple, publicly accessible threat hunting playbooks and using simple searches and statistical methods such as frequency distribution analysis can result in successful hunts. Even less experienced analysts can perform these hunts.

“Threat hunting can be fully automated.”
The analytical nature of the threat hunting process prevents it from being fully automated. Humans will always be a necessary part of the process. Data collection and enrichment steps, however, often can be automated.

“Threat hunting always finds attackers or malware.”
In the majority of cases, threat hunting will find policy violations and misconfigurations more than often than malware or human attacker activity. These detections are also highly valuable and actionable for any organization, and addressing them will result in a baseline with less noise, which makes threat hunting easier. Nevertheless, threat hunting will eventually find evil, especially in the early stages of the program when an organization first performs a few hunts. Because results are actionable and might trigger the incident response process, analysts tend to joke about why it is a horrible idea to start threat hunting on a Friday afternoon.
The SANS whitepaper “The Who, What, Where, When, Why and How of Effective Threat Hunting” explains the threat hunting process details at length. In this paper, however, we provide a high-level summary of threat hunting, as presented in Figure 1.

**Before You Begin**

If you are just getting started with your threat hunting program, keep these considerations in mind.

1. First and foremost, you need to have time and personnel dedicated to threat hunting. The most efficient threat hunting programs have analysts’ time explicitly assigned for hunting activities.

2. Decide who will do the hunting. When you choose the organizational model for your hunt team, think about your organization’s structure, the size of your company and your budget. Ideally, you need more than one analyst, and the best threat hunting teams have members with a diverse set of skills.

3. Focus on high-impact malicious activities first! The easiest way is to start with intelligence-driven hunting. Check out other threat hunters’ playbooks to jump-start your program.

**Form a hypothesis (also known as threat hunting lead).** Look for a known attack step, attack phase or artifacts generated by malware or a threat actor. The hunt team determines which data sources are required for the analysis and which is needed to test the hypothesis.

**What to search**—Threat hunters choose the best way to display that data: single search, a more elaborate statistical method or a visualization technique.

**How to search**—Collect, enrich, manipulate and analyze data by looking for fields and anomalies that might contain evidence of attacks.

**How to focus**—Transform a successful threat hunt into a new detection method or at least automate filtering (whitelisting) of known-good for the next threat hunt.

Threat hunting entails a more mature organization with a defensible network architecture, advanced incident response capabilities and security monitoring/security operations team. The SANS whitepaper “The Who, What, Where, When, Why and How of Effective Threat Hunting” mentions the Hunting Maturity Model (HMM) from David J. Bianco, which describes five different categories of an organization’s hunting capability.

---


A Word on Automation

Automation can save time and allows threat hunters to focus on new hunting scenarios instead of running the same ones repeatedly. As previously stated, however, we can never fully automate threat hunting. We can automate data retrieval and transformation steps most of the time, but the analysis and interpretation of data require human analysts. Machine learning solutions can help, but decision making based on the analysis still needs human judgement.

Google’s “Hunt Once” rule is a more advanced approach, which asks you to perform a threat hunting scenario once and then try to automate it as much as possible. This approach requires:

- Very skilled analysts with development experience or at least scripting skills
- Robust, specialized datasets and tools
- A security analyst group that can tolerate a higher amount of false positives

Organizational Models

There are various structures for threat hunting teams, and they depend more on an organization’s culture and structure than on the size of the security team. The most common models are shown in Table 1.

Metrics

Because threat hunting is very labor- and time-intensive, CISOs need to justify the resources (time, personnel, infrastructure and tooling) allocated to it. Carefully selected and presented metrics can effectively communicate the progress made in your threat hunting program and prove you have implemented it correctly.

As mentioned in the “SANS 2018 Threat Hunting Survey Results,” a threat hunting program can improve many domains in information security, and teams should measure improvement to show the maximum

---

Table 1. Threat Hunting Team Structures

<table>
<thead>
<tr>
<th>Team Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>External Hunt Team</td>
<td>This model essentially outsources your threat hunting activity and can introduce your organization to the concept and benefits of threat hunting. Nonetheless, an external team will never understand your environment like you do, and as a result, this model is less desirable for the long term.</td>
</tr>
<tr>
<td>Dedicated Hunt Team</td>
<td>Typically in large organizations and government entities, a small team of skilled, full-time employees threat hunts. They might have to help with alerts in the Security Operations Center (SOC), but they are not spending most of their time with alert analysis. The drawback of this structure is creating silos between SOC analysts and hunt team members.</td>
</tr>
<tr>
<td>Combined/Hybrid Team</td>
<td>A team member might have a combined role of SOC analyst/threat hunter, incident responder/threat hunter or security team member/threat hunter doing daily hunting in addition to their other duties. SOC analysts already have most of the skills required for threat hunting, so this is an obvious step forward for organizations. This is a typical model for smaller organizations or smaller teams with skilled analysts and strong detection capabilities. The risk in a combined role model is analysts might have other priorities, leaving no time for threat hunting at the end of the day.</td>
</tr>
<tr>
<td>Periodic Hunt Teams</td>
<td>In this model, security team members are periodically pulled away from other work to form a threat hunting team. This might happen weekly, bi-weekly or monthly and it requires a clear plan with a specific task to be effective. This model works for large organizations with a large pool of SOC analysts as well as for smaller ones with a small security team, performing threat hunting just for a few hours per week. It is important to rotate team members to give everyone exposure to threat hunting.</td>
</tr>
</tbody>
</table>

---

return on investment. Following the categorization introduced in “PRAGMATIC Security Metrics,” we divided our examples into operational, tactical and strategic metrics, as shown in Table 2.

### Table 2. Threat Hunting Metrics Categories

<table>
<thead>
<tr>
<th>CATEGORY: Operational</th>
<th>WHO MEASURES: Threat hunters</th>
<th>AUDIENCE: Threat hunting managers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time spent by analysts with data types, searches, pivots and dashboards</strong>—This metric provides insight into the value of data types and analysis techniques.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data coverage and retention time</strong>—These metrics can direct attention to the next steps for data collection. Data must be available to ensure the success of a threat hunt. Measure the availability of different data types per network segment or organizational units (or even per endpoint granularity).</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data hit/miss ratio per data source</strong>—A hit is whenever you use a data source during a hunt, whereas a miss is when a needed a data source in a hunt was not available (maybe due to of a lack of coverage or the data retention time was too short).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CATEGORY: Tactical</th>
<th>WHO MEASURES: Threat hunters and their managers</th>
<th>AUDIENCE: Threat hunting managers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dwell time or mean time to detect (MTTD)</strong>—While threat hunting reduces MTTD in theory (it finds threats all other security controls missed), it can be a long time before you find the actual threat. Therefore, MTTD does not exactly express the efficiency of threat hunting.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hunting efforts mapped to models such as the MITRE ATT&amp;CK Matrix or the Cyber Kill Chain</strong>—Use any coverage such as the number of hunts, the number of detections or the analysis time spent on each element of the matrix or the kill chain.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CATEGORY: Strategic</th>
<th>WHO MEASURES: Threat hunting managers</th>
<th>AUDIENCE: Executives and threat hunting managers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of successful hunts</strong>—Every incident detected is a win for an organization; we perform threat hunting because other detections failed. The number of successful hunts can be broken down into further categories based on the issues discovered (policy violations, malware, human attackers). Executive reports on successful threat hunts can also monetize the loss prevented by mapping to Ponemon Institute’s Cost of a Data Breach studies.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Data Sources for Threat Hunting

Threat hunting requires data types similar to those you likely use for alert analysis, but SANS has a higher preference for detailed logs compared to alert data or full packet capture. You can divide data sources into network data and endpoint data, but you also need threat intelligence data and information about the environment to perform

---

12 “MITRE ATT&CK™,” https://attack.mitre.org
analysis. Getting access to network data is typically easier than getting endpoint data, as the former is usually already present on a network device (such as proxy logs or firewall logs). Figure 2 illustrates the different data types required for threat hunting.

In addition to network and endpoint data sources, data enrichment techniques can add new data fields like reputation scores you could use in conjunction with the already existing data fields. You might be able to use some added fields such as Autonomous System Numbers (ASNs) to find evidence of potentially unwanted activity. These added fields can be:

- Calculated values such as frequency or entropy values
- External lookups such as reputation scores, geolocation information or ASN lookups and domain registration information
- Results of data field manipulation, such as separating the top level domain (TLD) and subdomain in case of a domain name or separating the path and filename case of an absolute path

Tools allowing hunters to perform various data transformations and manipulations are crucial to performing a proper analysis. These tools include simple searches and correlations as well as pivoting between different data types and dashboards. Visualizations and statistics displaying the change of the value of specific fields over time are also vital for investigations. One of the most often used threat hunting techniques is frequency distribution (also known as stacking, aggregation or grouping); it simply shows the number of instances of various outcomes in a sample.

Figure 3 presents a simple example of frequency distribution. The tshark command extracts the HTTP User-Agent strings from the sample PCAP file, and the values are counted and ordered by the number of occurrences in ascending order. This technique allows an analyst to find outliers, such as the two values marked at the top. Other User-Agent values in the list would require further investigation in a typical enterprise environment.
Starting with network data might be the most manageable approach. Some of the more important network data sources include those in the following list.

- **Proxy logs** (or some other form of HTTP transaction logs) are rich data sources in an organization. Threat hunters can find traces of human web browsing activity as well as traffic from automated tools. Proxy logs have important fields such as URL requested, HTTP method, HTTP status code, request size and in some cases, the query string, user-agent strings and Referer URLs.

- **DNS logs** are another essential network data source. Ideally, they should include the DNS query and the response too, but many organizations only log DNS queries. In either case, you have the domain requested, the system making the request and the requested record type. DNS logs can be used to find artifacts of a wide variety of malicious activity such as fast-flux, DNS rebinding, phased command and control (C2 or C&C) and domain name generation algorithms.

- The format and the fields existing in **firewall logs** depend on the given firewall platform, but they mostly contain layer 3 (IP addresses) and layer 4 (ports, state) information. They still can be very valuable (for example, when you are looking for known malicious traffic going to a C2 server or searching for signs of beaconing activity).

- Similar to firewall logs, **NetFlow records** provide a summary record of communications with layer 3 and layer 4 information. NetFlow is ideal for getting statistical information about the network traffic, and you can use it to create aggregations such as top talkers based on the number of bytes, number of packets or the duration of the connections.
• Today’s Internet traffic is mostly encrypted, but SSL/TLS logs and X.509 certificate logs still have a few precious plaintext fields you can use during threat hunting. You can further enrich this data with information such as JA3 and JA3S hashes15 (invented by Salesforce security engineers) to fingerprint SSL/TLS clients and servers.

• There might be several other network data types in your organization, such as access logs to VPN and cloud services or other service logs like SSH or even FTP. The significance of these service logs depends on the given environment.

• Full packet capture is a network data type used less often during threat hunting, simply because not every environment can afford to have it and because the sheer volume of it makes it challenging to handle and analyze.

Endpoint Data

On the endpoint data side, you need some means to access or collect the data from multiple systems. It typically requires configuration changes and deployment of agents or other client software. The most common endpoint data sources used for threat hunting include those in the following list.

• The most obvious category is system event logs (Windows or Linux/UNIX). These logs cover a wide range of events, including some of the more interesting ones such as failed and successful login and logoff events, user creation events, group membership changes and PowerShell events.

• Along with standard system logs, Windows or Linux/UNIX operating systems have various solutions to generate extended system logs. For Windows systems, it is possible to turn on command line logging for process creation events or use the free Sysinternals Sysmon to generate additional logs for events such as service creation events or network connection events. On the Linux/UNIX side, auditd can be used to write detailed audit records to log files.

• Windows Registry keys are fundamental for Windows systems and contain a plethora of information you can use for threat hunting. Look for the “run keys” in the Registry, which are frequently used by malware and attackers to maintain persistence through system reboots.

• Application server logs such as web server logs, email transaction logs and database logs are also excellent sources for threat hunting.

• Similar to but less common than full packet capture, system memory is a threat hunting data source that it is harder to obtain and analyze. Sometimes, however, it is the only source of artifacts for more sophisticated attacks.

---

To formulate your threat hunting lead or hypothesis, rely most often on some form of threat intelligence, as mentioned in "Thinking like a Hunter: Implementing a Threat Hunting Program." On a more operational threat intelligence level, we most often use some sort of indicators of compromise (IoC) or reputation data. Sources of tactical and strategic threat intelligence can be industry- or company-specific reports, information from previous incidents or maybe internal or external red team assessments.

Some level of information about the environment is indispensable for successful threat hunting. The hunting team should have or be able to get information about internal systems, assets and network segments. Get a list of approved and installed software in your organization and be able to query the attributes of users and groups. More mature hunt teams eventually develop a deeper understanding of normal user behavior in their organization, such as standard working hours and typical user login activity to systems.

**Tooling Considerations**

To support your threat hunting activities in your organization, you have to take into account the tooling needs of your team. These tools are not necessarily specific to threat hunting and most environments probably already have some of them, but they have to be used slightly differently for hunting.

These tools fall into the following main categories:

- **Log aggregators and security information and event management (SIEM) solutions**—Threat hunting requires access to several different data types, and the centralized collection of these various data sources allows analysts to perform enterprise-wide correlations and searches. Visualization features of these platforms allow threat hunters to find signs of compromise and display relationships in the data that would be hard to find otherwise. SIEM platforms can serve as a threat hunting collection and analysis tool, however, you only need some of the SIEM functionality for threat hunting. You will not need SIEM features such as alerting or reporting.

- **Analyst supporting tools**—Depending on the threat hunting scenario, these can be highly specialized analysis tools, like the ones used for analyzing full packet captures or simple shell tools and scripting languages to develop custom methods of detecting specific threat actors.

- **Endpoint data collection tools**—Typically endpoint detection and response (EDR) tools fall into this category. They allow analysts to query a large number of endpoints across the organization relatively fast. There are free and open source software (FOSS) solutions for endpoint data collection, and based on the SANS 2018 Threat Hunting Survey Results, there are also a few commercial platforms addressing the same problem.

---


In addition to proper tooling, threat hunting activities require creating one or more test
environments. Increasing the log level to the maximum in a test lab allows the analysts
to perform attacks, run malware and different tools, and look for the artifacts generated.
The information gathered from these test systems then can be used to create leads
and hypotheses, and threat hunters can start looking for these artifacts. One of better
known, ready-to-go test environments is Detection Lab from Chris Long.18

For threat hunting teams, a knowledge base is even more important than tools or test
labs. It can be a simple wiki or similar collaboration platform allowing information
sharing among team members. Analysts should make detailed notes during threat
hunting and later turn them into knowledge base entries. These organized, searchable
entries will be threat hunting playbooks containing searches, data transformations,
and scripts with observations (references to users, applications, systems in the
environment), detected anomalies and links to external resources. These entries can be
used by other team members to learn about and understand previous hunts; therefore,
the descriptions should refer to known models such as the Cyber Kill Chain, the
Diamond Model or the ATT&CK Matrix.

**Conclusion**

Threat hunting is more than just a popular expression
among vendors. It is a powerful, proactive detection
method. A relatively mature organization can start a threat
hunting program by planning and allocating time and
people, but the undertaking does not call for expensive
tools or years of experience. Starting small and gradually
extending on the data types and scenarios is key to a
successful program. Remember even if your threat hunting
does not always find signs of compromise, it dramatically
increases your visibility and understanding of your
environment.

**How to Start a Threat Hunting Program**

1. Allocate time and personnel dedicated to threat hunting.
2. Choose an organizational model for your hunt team.
3. Choose the right tools for the job and make sure you collect
   the data you need.
4. Focus on high-impact, malicious activities first and follow the
   threat hunting process:
   a. Form a threat hunting hypothesis!
   b. Decide what to data sources to search.
   c. Determine how to search (searching and visualization).
   d. Focus your hunt on specific data fields and data
      manipulations.
   e. Ask yourself whether you can automate some of the steps.
5. Measure and communicate your results using metrics!

---

**About the Author**

David Szili is a SANS instructor of SANS FOR572: Advanced Network Forensics: Threat Hunting, Analysis, and Incident Response, as well as a managing partner and CTO at a Luxembourg-based consulting company. He has more than eight years of professional experience in penetration testing, red teaming, vulnerability assessment, vulnerability management, security monitoring, security architecture design, incident response, digital forensics and software development. David holds several IT security certifications such as GSEC, GCED, GCIA, GCIH, GMON, GNFA, GYPC, GMOB, OSCP, OSWP and CEH, and he is a member of the BSides Luxembourg conference organizing team.

---

**Sponsor**

SANS would like to thank this paper’s sponsor:

![CISCO](image-url)
## Upcoming SANS Training

Click here to view a list of all SANS Courses

<table>
<thead>
<tr>
<th>Training Title</th>
<th>Location</th>
<th>Date Range</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SANS San Jose 2019</td>
<td>San Jose, CAUS</td>
<td>Aug 12, 2019 - Aug 17, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Prague August 2019</td>
<td>Prague, CZ</td>
<td>Aug 12, 2019 - Aug 17, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Minneapolis 2019</td>
<td>Minneapolis, MNUS</td>
<td>Aug 12, 2019 - Aug 17, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Amsterdam August 2019</td>
<td>Amsterdam, NL</td>
<td>Aug 19, 2019 - Aug 24, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Virginia Beach 2019</td>
<td>Virginia Beach, VAUS</td>
<td>Aug 19, 2019 - Aug 30, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Hyderabad 2019</td>
<td>Hyderabad, IN</td>
<td>Aug 26, 2019 - Aug 31, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Philippines 2019</td>
<td>Manila, PH</td>
<td>Sep 02, 2019 - Sep 07, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Munich September 2019</td>
<td>Munich, DE</td>
<td>Sep 02, 2019 - Sep 07, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Brussels September 2019</td>
<td>Brussels, BE</td>
<td>Sep 02, 2019 - Sep 07, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Canberra Spring 2019</td>
<td>Canberra, AU</td>
<td>Sep 02, 2019 - Sep 21, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Network Security 2019</td>
<td>Las Vegas, NVUS</td>
<td>Sep 09, 2019 - Sep 16, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Oslo September 2019</td>
<td>Oslo, NO</td>
<td>Sep 09, 2019 - Sep 14, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Dubai September 2019</td>
<td>Dubai, AE</td>
<td>Sep 14, 2019 - Sep 19, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Raleigh 2019</td>
<td>Raleigh, NCUS</td>
<td>Sep 16, 2019 - Sep 21, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>Oil &amp; Gas Cybersecurity Summit &amp; Training 2019</td>
<td>Houston, TXUS</td>
<td>Sep 16, 2019 - Sep 22, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Rome September 2019</td>
<td>Rome, IT</td>
<td>Sep 16, 2019 - Sep 21, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Bahrain September 2019</td>
<td>Manama, BH</td>
<td>Sep 21, 2019 - Sep 26, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS San Francisco Fall 2019</td>
<td>San Francisco, CAUS</td>
<td>Sep 23, 2019 - Sep 28, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Dallas Fall 2019</td>
<td>Dallas, TXUS</td>
<td>Sep 23, 2019 - Sep 28, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Kuwait September 2019</td>
<td>Salmiya, KW</td>
<td>Sep 28, 2019 - Oct 03, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Northern VA Fall- Reston 2019</td>
<td>Reston, VAUS</td>
<td>Sep 30, 2019 - Oct 05, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS Cardiff September 2019</td>
<td>Cardiff, GB</td>
<td>Sep 30, 2019 - Oct 05, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>Supply Chain Cybersecurity Summit &amp; Training 2019</td>
<td>OnlineVAUS</td>
<td>Aug 12, 2019 - Aug 19, 2019</td>
<td>Live Event</td>
</tr>
<tr>
<td>SANS OnDemand</td>
<td>Books &amp; MP3s OnlyUS</td>
<td>Anytime</td>
<td>Self Paced</td>
</tr>
</tbody>
</table>