Mandiant

ACQUIRING DATA FROM A THREAT HUNT BY USING INTELLIGENCE

This article focuses on the cyber defense function: **Threat Hunting** – using intelligence to frame queries to find evidence of advanced persistent threat (APT) groups within your environment.

Read Time: 10-12 Min

If you haven't read our primer on the importance of the Threat Hunt Mission, we recommend you spend 6-8 minutes reviewing the article **Introduction to an Intelligence-led Threat Hunting Framework (https://docs.mandiant.com/home/introduction-to-an-intelligence-led-threat-hunting-framework)**.

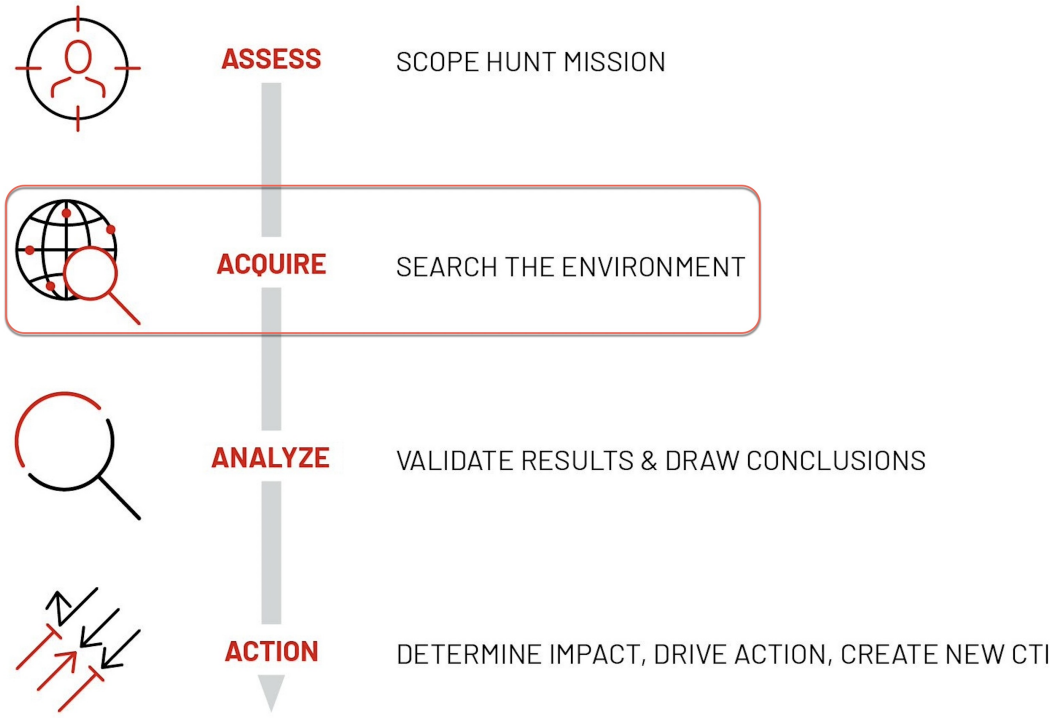
To recap, Mandiant defines Threat Hunting as **the methodical, use-case driven, proactive identification of cyber threats within a computing environment or infrastructure**.

Objectives and Key Actions for Application

Cyber threat intelligence is required to help identify threats we should be looking for, based on the likelihood and impact of those threats. In other words, organizations should be actively looking for those threats they expect to target them.

Threat Hunting complements security detection. Organizations often focus on reactive defense - identifying the presence of attacks through alert generation of their malicious activities by deployed security monitoring solutions. Threat Hunting is proactive – actively searching through your environment to detect and isolate signs of malicious activity that have evaded existing security monitoring solutions. These two processes complement one another.

For this use case article, we'll be focusing on how to frame a Threat Hunt Query as part of the Acquire phase of Mandiant's A4 Threat Hunting Framework: How to Effectively Search the Environment for Evidence of Threat Activity.



For a more detailed dive into the A4 Threat Hunting Framework, see **Introduction to an Intelligence-led Threat Hunting Framework (https://docs.mandiant.com/home/introduction-to-an-intelligence-led-threat-hunting-framework)**.

Approaching the role of data acquisition

The role of data acquisition appears self-explanatory: once you've scoped what/whom you are searching for and why, then searching the environment for evidence of their presence or attempted entry seems relatively straightforward.

However, like scoping, those who activate searches across production environments need to understand what they are seeking, what they might find, and the broader implications of interrogating large systems and datasets upon business operations. For each of these components, those who activate the search will need to develop clear criteria on the type of data they may encounter, how to handle this data, and how much data is required to disprove the Threat Hunt hypothesis.

The volume of data is likely going to have a direct correlation with how complex the search becomes. Too much data within search results may over-complicate the Analyze phase, returning traffic results from a large portion of the network that over-tax the capability to effectively find meaningful results. Too little data may similarly lead to inconclusive analysis that doesn't progress the understanding of the threat actor under investigation. To find the balance of enough data to inform but avoid overwhelming analysts, we recommend searches be designed to scale up over time. The initial search should be contained yet thorough – replete with enough data points to render useful analysis yet applied against a small, identifiable subset of the network. Once the search parameters have been confirmed as valid, threat hunters can then expand their reach, broadening their reach into the network.

Within this method is an important implication: threat hunting will take time and the success of the results will be commensurate with how often the search is updated and refined. Not finding anything isn't proof nothing is there to find. Our recommendation regarding searching the environment is to iterate each search against specific threats until either (a) the results are comprehensive and as definitive as possible, or (b) the threat declines in importance and positive results contain less priority compared with current threats.

As always, the Cyber Threat Profile will ultimately guide the priorities of threats. To help align your Threat Hunt Missions to the broader intelligence organization's Cyber Threat Profile (CTP), spend time to read our article on **Creating a Cyber Threat Profile (https://docs.mandiant.com/home/creating-a-cyber-threat-profile)**.

How To Effectively Search Your Environment

Like scoping the threat query, effectively searching your environment will rely on your capability to establish clear processes and procedures that allow for repeat performances while being able to account for changes in variables. (For more information about scoping the threat query, see **Scoping a Threat Hunt with Intelligence (https://docs.mandiant.com/home/scoping-a-threat-hunt-with-intelligence)**.)

Visibility

While many security operators are adept at identifying sources of data or information available within their network, understanding the reverse situation is just as important: what data or information is not available.

There are many reasons why data may not be available to threat hunters, ranging from issues with system permissions, data being siloed, sensitive systems, misconfigured applications that obstruct or prevent access, or simple reluctance on the part of coworkers to share access.

Decisions will need to be made about rectifying difficult access constraints. There is a balance between how hard access is to obtain and the criticality of the data. Obviously, the more vital the data, the more effort should be made to overcome access challenges.

Regardless, threat hunters will need to maintain an accurate record of what gaps exist within their coverage to understand their degree of visibility. Significant gaps will need to be reconciled within the Analyze phase, as a noted Assumption, or within a Caveat, or both.

Scoping was about defining the boundaries for the impending Threat Hunt activity. If the parameters were too broad or too narrow, the search would be ineffective. The process for searching - or interrogating your environment - needs to adhere and support this scoping. This means the data acquisition phase of the A4 Threat Hunt Framework focuses on identifying the tools and anticipating the type of data these tools would likely collect against the parameters set by the scope. To ensure the search is completed effectively, you should give active thought to (a) the types of data that will likely be returned based on the system interrogated and tool used, and (b) the type of data that is needed to best analyze the presence of the threat actor outlined in the scope.

Finally, sufficient thought must be given to the impact of the data on the system. Data acquisition needs to capture the impact, influence, and timing of the data within or upon the environment.) For example, how sensitive and accessible is the data? Is it being retained longer than necessary? What would happen if it was manipulated or exfiltrated?) Changes to the fidelity of the data should be minimized (or at least, noted) to ensure the reliability of search results are not skewed.

Environment interrogation process for Threat Hunting (TH)

The following steps provide a thorough breakdown of considerations when interrogating your environment. Click on each for more details.

1. Consult the scoping

Familiarization with the scope (or parameters) of the proposed threat hunt is critical. The scope, preferably within a template, contains the ‘terms of reference’ that significantly impact the interrogation of the environment. All the remaining steps within environment interrogation need to align with the scope – "out of scope" is often used as the justification to limit or restrict complex tasks or projects from becoming unwieldy and ineffective.

Likewise, scope-creep is a common problem whereby tasks or projects start to expand due to the perception that all problems encountered need to be solved. Undoubtedly, in executing the search for data and behavioral evidence, there will likely be decisions to be made to include or exclude certain parameters. Analysts should remain flexible to account for the messiness of any intelligence process when put into practice. However, conflating threat hunt missions runs the danger of returning confusing datasets and search results which complicate analysis. Ideally, indications of other threat activity adjacent or unrelated to the scope should be flagged for handling in a separate process to ensure the fidelity of the Threat Hunt originally scoped. The scope should be adjusted in subsequent iterations of the same Threat Hunt - altering search parameters mid-stream can return "tainted" results. The scope is there to keep data collectors and analysts honest - use it to keep yourself on rails and locked into one area of focus at a time.

Follow Through

There is a temptation to adjust searches mid-stream when zero or little results have been returned. Don’t rush to alter the search parameters and run the search again. Little or no results could be the result of either

poorly conceived search parameters or well-constructed search parameters! Perhaps what exists wasn't found, or perhaps there was nothing to find. It can be difficult to tell in the moment.

The way to differentiate these scenarios is through consistent application of the same, entire process over time. Nothing invalidates search results quicker than someone having to ask, "Did we finish this search?"

See the A4 Framework through, even if some phases may be truncated. Knowing what you've tried is an invaluable input for future iterations.

2. Define searchables

A searchable (or entity) is a type of data that can be...searched for! Defining searchables is like figuring out what type of physical or behavioral evidence may exist to prove a crime. In some cases, we can see the impact of the crime. However, to prove the crime took place (and to attribute the crime to a specific cluster or threat actor), we need to define what types of evidence exist that, when assessed together, may indicate malicious activity.

Defining searches for cyber activity is important as threat actors can often manipulate or co-opt legitimate system functions to aid in their malicious activity. Thus, based on the scope, defining the "hard" physical evidence and the behavioral, pattern, or profile evidence can assist in building a complete picture of the type of data and information we should be seeking within the environment.

TTPs

TTP stands for tactics, techniques, and procedures. They are the characteristics that define the way a threat actor prefers to operate. In other words, TTPs represent the digital fingerprints of threat actors.

TTPs aren't hard and fast - threat actors will change them to suit their purpose and to remain current with new technologies. But they can provide a comparison point to help in building searches.

A full list of searchables appears at the end of this article. Be creative when building searches - malicious activity is often indicated by the use of a collective set of tools used in a certain combination to achieve a specific outcome.

3. Identify access and tool requirements

Obviously, when you need to interrogate the environment, you'll need to understand where and how to search. Once you've identified the internal tools you can use, you'll need to acquire access and permission to areas of the network that would benefit from being included within the search parameters. This will necessarily involve data stakeholders who act as custodians of the network and relevant systems. Keep them apprised of your intentions, including exact dates and times when you'll be executing searches. You don't want to waste their time and resources investigating your activity as possibly malicious behavior.

Involve stakeholders and custodians early and ensure they are treated as interested parties – after all, they are! The results will be of direct benefit to their security requirements and duty of care over data under their control. When seeking access and considering tools to use, ask data stakeholders for their advice: they are likely to best understand the nuances of the network, systems, and data under their watch.

4. Prepare to aggregate and process results

Be prepared to accept and process the quantity of data expected. Search results are unlikely to be clean and neat, or sorted into "like-for-like" categories with no outliers. Rather than trying to anticipate the exact degree of messiness in the search results, it is better to build an understanding of how best to aggregate and process the data based on the parameters of the search you intend to run. These parameters should reflect the nature and type of threat actor you're searching for: you can anticipate the type of data and searchables that would indicate malicious presence or activity of the threat actor(s), and prepare to aggregate and process around those factors.

If we think about this stage in terms of crime, the nature and type of crime would dictate the type of evidence we're interested in collecting. Homicides and assaults would rely on physical evidence that needs to be handled in such a way as to preserve the DNA or physical specimens. Aggregation and processing of this evidence would focus on avoiding contamination and obeying the chain of custody. Other crimes may have lowered burdens of proof and thus rely on other types of evidence. In all cases, the nature of the crime determines what evidence to expect and how to handle it. Cyber crime is no different: different malicious cyber activity will require acquiring different types of data. Preparing how best to handle, preserve, and process this data is important for efficient analysis.

5. Initiate data collection

If this is the first iteration of a new search, start small. Look to test queries and search syntax against a subset of data where the potential results are limited or can be easily analyzed. This ensures that any mistakes, search failures, or unexpected results can be more easily identified, verified, and amended. Successful or efficient searches can be expanded in future iterations to encompass more data and systems, enabling you to become proficient with the expanded results.

Don't ignore failures, and don't try to apply fixes without proper documentation. The documentation doesn't need to be extensive, but documenting the failures of syntax or queries can preserve this knowledge for future iterations or for those later seeking to expand, improve, or base their query on the original iteration. In addition, repeat failures can demonstrate configuration or control issues within the network or system that can be rectified.

Likewise, searches may provide information on data unavailability, parsing errors, or performance issues. Documenting these issues during the data collection phase will improve your process and may be used to justify the further allocation or acquisition of resources and tools.

6. Validate results

Validation of results will depend on what you expected to have returned based on your query construction and scope. These expectations will also be a comparison between the data points returned and those expected based on the query logic.

This is another stage to determine whether there are any parsing or query issues. Note that this isn't the time to analyze the results or to rebuild and re-run a similar query. As noted above, finishing each query will ensure the process remains consistent with each search. Failure to complete the entire process with each search will bring results into question. This step is simply a check to ensure you're getting the type of results you expected, based on quality and volume of data returned.

7. Perform initial analysis

Two stages within the Acquire phase of the A4 Threat Hunting Framework are devoted to analysis: validation of results and initial analysis. But how do activities associated with these stages differ from the Analyze phase?

Validation and initial analysis occur within the collection phase of the Intelligence Process (or lifecycle). From a process perspective, collectors are not analysts. Collectors collect raw data; analysts examine the data to provide meaning. As a

best practice, ensure these functions are separate.

However, that's not to say that collectors don't analyze the data they collect. They must ensure that the data they provide to analysts is useful. This includes identifying where gaps in data collection exist, and providing context that may influence how the analysts evaluate the data. Thus, collectors analyze data to ensure the data they collect is relevant, processed, clean, and provided in a format that allows it to be manipulated by analysts.

Experienced collectors may be able to add greater context regarding the data collected, based on similar results in the past or nuances within the data itself, storage methods, or the way data is collected. For instance, if collectors are aware that log data is only kept for 6 weeks, but the search parameters specify collection over 8 weeks, collectors need to either shape their query to fit or note the discrepancy so it's taken into account when examining results.

Data Fidelity versus Processing

Within cyber security, a tension exists between preserving the data as collected and cleaning and processing data for it to be useable and human-readable. Where possible, data should be collated and consolidated into similar attributes ("like-for-like"). This helps analysts to compare similar datasets, among other activities.

Limits must be applied to data processing, however. Processing shouldn't eradicate or change unique attributes within the data or filter the data to the degree that loses important attributes. Not all data can be made similar.

A good rule of thumb when processing data is to ensure the data's fidelity isn't over-powered by any cleaning done to ease the burden of the analyst. Preserve the data's format and originality first.

8. Determine priority of results and escalate critical results

Part of the initial analysis phase is to determine the priority of results, especially if query results indicate a significant malicious presence, or the presence of sophisticated threat activity.

Prioritizing results can be based on a few different parameters: the volume of positive results (indicating a potentially significant malicious presence), the determined criticality of the search itself (as determined through the scope phase), and the likely impact to the organization suggested by the results.

Experienced collectors or those that construct and run searches are also often in a good position to apply the "sniff test" – if an often-run search returns different results than usual, this could be an indication that something has changed or that results warrant further analysis sooner rather than later. Don't ignore results that seem somehow "off" or unexpected. It may be a harmless quirk, but it could be something significant.

Ideally, you should have a specific escalation process that prioritizes critical results or results that indicate a critical incident. As above, these can be based on either the unexpectedly high volume of positive results, or the positive return of results that indicate a significant impact on the network. These results should be raised to a managerial resource for evaluation and then fast-tracked into the Analysis phase for further verification and meaning.

9. Document findings and record processes

Cyber threat intelligence works best when activities can be repeated in a consistent manner with past iterations informing future iterations. Repeated and consistent application of intelligence activities reduces and mitigates errors in analysis

caused by personal biases and differing methods of collecting the same information over time.

Defining and documenting processes and findings is one of many ways to aim for consistency. It's especially important when intelligence activities, such as threat hunting, involve repeated tasks with minimal changes between iterations. Identifying processes - a set way to perform certain tasks efficiently - can aid in training inexperienced staff and ensuring tasks are performed in a similar fashion no matter the person involved.

Documenting findings also ensures the capture of a body of contextual knowledge over time regarding similar tasks and activities. Decisions can be made more efficiently knowing what has been tried, failed, and succeeded in the past. While documenting findings helps to avoid unnecessary duplication of effort, its primary function is to justify further intelligence activity downstream. Why some search results were prioritized over others may be self-evident in the moment, but memories fade fast. Written documentation is essential to justify actions and decisions made, particularly if security audits are required in the future.

10. Consider frequency and automation

Documenting findings and recording processes both seek to make threat hunting more efficient by making life easier for those who follow. Over time, conclusions and patterns can be identified for repetitive tasks that may benefit from automation.

Generally, threat hunt queries will need to be re-iterated and run frequently and/or regularly. A single threat hunt query is essentially a snapshot, which has limited benefits for an active environment and an ever-changing threat landscape. Designing a threat hunt process is important: be able to reproduce same or similar queries with minimal resources and effort. The next step beyond a process is automation.

Establishing an automated process to run threat queries has obvious advantages, not the least of which being able to free up finite resources to construct other queries to lessen your organization's risk exposure. Automation should be achieved using the least complex and most viable option – you don't want to waste the resources freed by automation by forcing them to attend to overly complicated automated scripts or programs. Automation solutions should be built to anticipate and accommodate upcoming changes to the environment, while never relying on one individual's knowledge or skillset. Generally, the less complex the automation, the easier it is to fix if something goes wrong.

Cyber Security as Deterrence

In general, threat actors can be divided into three broad camps:

1. **Opportunistic**: Those who run innumerable attack methods against a large target base ("scatter-gun approach")
2. **Targeted**: Those who conduct attacks with specific objectives and array against a narrow group of potential targets
3. **Dedicated**: Those who dedicate themselves to a specific target that holds unique, valuable, or sensitive data that has intrinsic meaning to the threat actor

In the first two cases, robust cyber security (protection, detection, mitigation) can help deter threat actors and push them towards other targets. The harder the attack surface is to penetrate, the more the threat actors will be encouraged to seek easier entry points at other targets. Threat actors won't want to expose their methods to detection without achieving rewards.

Sadly, the third group won't be as easily dissuaded.

List of Searchables

Following are a sample of the different types of searchables you may want to use to build searches and conduct analysis. The list is neither specific nor exhaustive - as you become more familiar with the Threat Hunt process and run more searches, you'll be able to define more specific and detailed searchables.

Searchables are going to depend on how threat actor(s) engage and influence their victims. Look for what they are likely to do, not what you would have done. Threat actors are creative, roguish, and human. The way they behave may not always be optimal or predictable.

- **Atomic IOCs**
  - These are static indicators, such as domains or endpoints of known actors.
  - While they are the most commonly thought-of type of searchable, they are innumerable and can age quickly into obsolescence.
  - The ability to alter them easily and quickly (i.e., domain registration) is a job for the Analysis phase.
- **Related infrastructure**
  - Pivoting across unique or exclusive infrastructure can help build an understanding of how threat actors are using the resources at their disposal.
  - As infrastructure is often the "bridge" between your environment and the threat actor's control, interrogating infrastructure is useful to determine your risk exposure.
- **Forensic artifacts**
  - Use Cyber Threat Intelligence to acquire information on the TTPs of attacker activity, such as infection vectors, malware and tools used, and how they move through the environment.
  - Remember, the indicator for malicious activity is often not just one tool in isolation, but rather the collective set of tools and how they interact with one another to affect the target.
- **Behavior and patterns**
  - Patterns can be used as searchables, such as a series of activities combined with identified artifacts to produce effects that achieve actors' goals.
  - Behavioral patterns may emerge during analysis. For example, social engineering that re-uses similar lures or campaign structure, exploits that use the same file extensions, or shared naming conventions for registered domains.
- **Target profile**
  - A target profile can be used as a searchable. Constructing a Cyber Threat Profile that details critical assets and threat actors likely to impact your organization can help narrow down parts of the network, or gateways through which threat actors will need to pass, to identify malicious activity
  - Understanding how your organization's network can be enumerated by an outsider through careful external reconnaissance can help form ideas regarding where initial access or attractive attack vectors to inform hunting queries.