Infrastructure Research and Hunting: Boiling the Domain Ocean

threatconnect.com/blog/infrastructure-research-hunting/

December 15, 2020

The Diamond Model of Intrusion Analysis identifies two main nodes as actor assets that may ultimately interact with a target / victim's own assets — capabilities and infrastructure. But while "exploitation" is usually considered something the adversary does, it works both ways as threat intelligence researchers and defenders in general can exploit the discoverable characteristics and tactics those adversaries employ with either node.

In this blog, we're going to explore important considerations and methodologies for exploiting actors' infrastructure tactics. Ultimately, our hope is to teach defenders how they can proactively find, address, and defend against their adversaries' procured infrastructure. Along the way, we'll provide examples of infrastructure research to serve as demonstrations of the concepts at play.

We're going to focus on adversaries' use of procured domains with this blog. In other words, most aspects of the workflows we describe herein usually will not be applicable to researching compromised infrastructure. From a procurement perspective, adversaries may take one or more of the following general steps to set up their infrastructure BEFORE it is used in operations:

- Create a registration persona
- Buy a domain name from a registrar/reseller
- Set up hosting at an IP address
- Set up target or operation-specific subdomain infrastructure
- Create an SSL certificate if requiring HTTPS communication
- Enable services at a hosting IP address or the domain
- Set up domain with a website or redirect

"Before" is the opportune word in the last sentence. Each one of those steps can leave behind a trace of the tactics that an actor used in conjunction with their procured infrastructure. In recent blogs, DomainTools described the inherent idea here that infrastructure indicators should be considered composite objects made of atomic parts that can be studied to extract adversary tactics. As defenders, if we can identify, research, and hunt for those characteristics and tactics, we can potentially identify our adversaries' infrastructure before it is used in operations against us. Ultimately, those steps manifest characteristics and tactics in one or more of the primary "dimensions" of infrastructure that we investigate:

- Registration (WHOIS)

- Hosting
- Subdomains
- Certificate use
- IP Configuration
- Content (HTML)

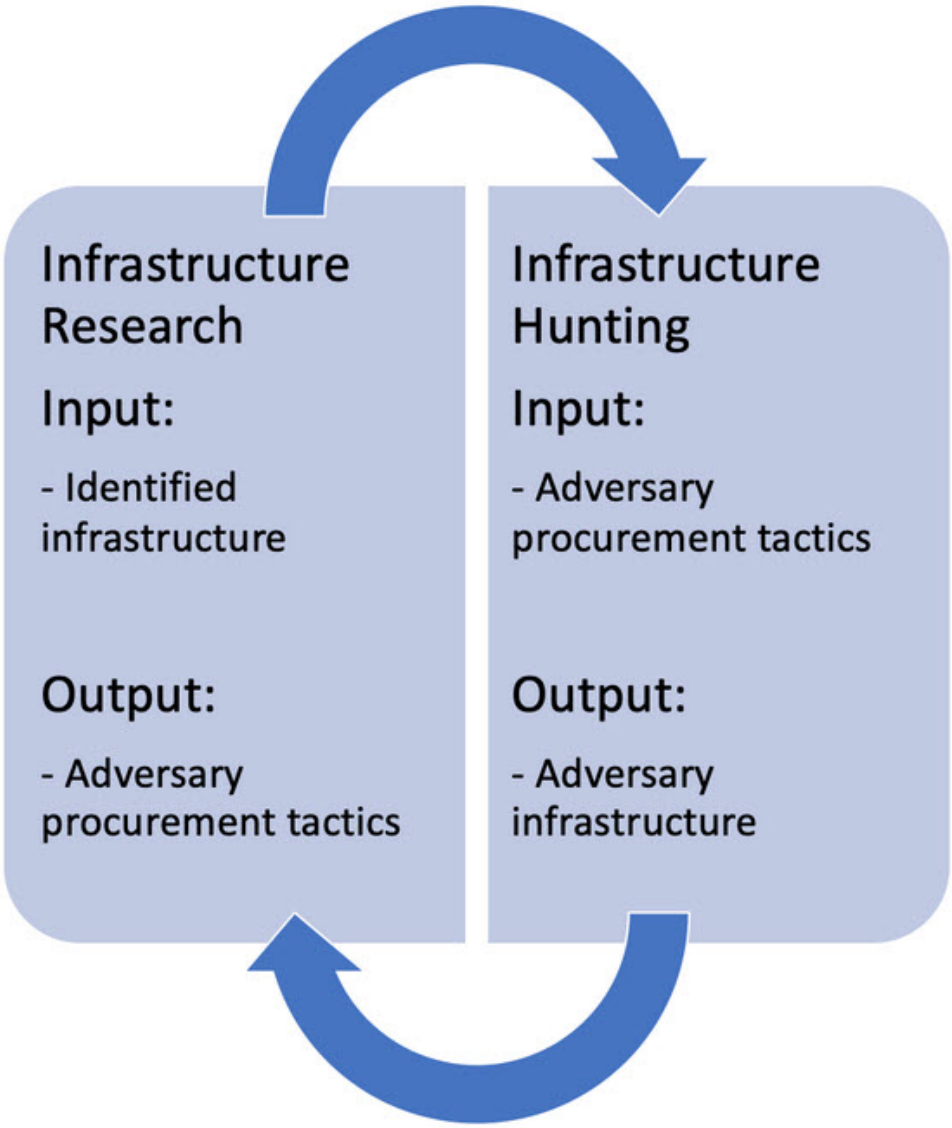
It's important to note up front that we're not just going to be looking for tactics that are exclusively unique to our adversaries. In some cases the tactics, or confluence of characteristics, that we hunt for may not yield adversary-specific results but rather a smaller set of infrastructure against which additional analysis may identify our adversaries' domains. All that is to say, even non-unique tactics can be exploited for viable hunting grounds if not used excessively.

Definitions

The terms "infrastructure research" and "infrastructure hunting" may be considered synonymous and used interchangeably to describe defenders' exploitation of adversaries' infrastructure characteristics (individual elements in a dimension) and tactics (combinations of characteristics). For the purposes of this blog, we will explicitly define each.

Infrastructure research refers to the retroactive efforts taken to understand the characteristics and tactics behind an adversary's identified infrastructure and building out an understanding of the adversary's current and past infrastructure based on those findings.

Infrastructure hunting refers to the proactive application of infrastructure research, where adversaries' known characteristics and tactics are exploited to identify their new infrastructure. There is a symbiotic relationship between the two, where output from each impacts the other.



Relationship between Infrastructure Research and Hunting

Tools

There are a number of tools that enable defenders to exploit adversaries' infrastructure tactics, in many cases overlapping with other capabilities. Each of the tools has its own strengths with respect to researching or hunting for adversary infrastructure, and in many cases using multiple in conjunction may be necessary to ultimately exploit adversary tactics. On the ThreatConnect Research team, we have found success using the following for research and/or hunting:

- DomainTools Iris
- PassiveTotal
- Farsight DNSDB and Scout

- urlscan.io
- Censys
- Shodan
- GreyNoise

Relationship to the Capabilities Node / Malware Hunting

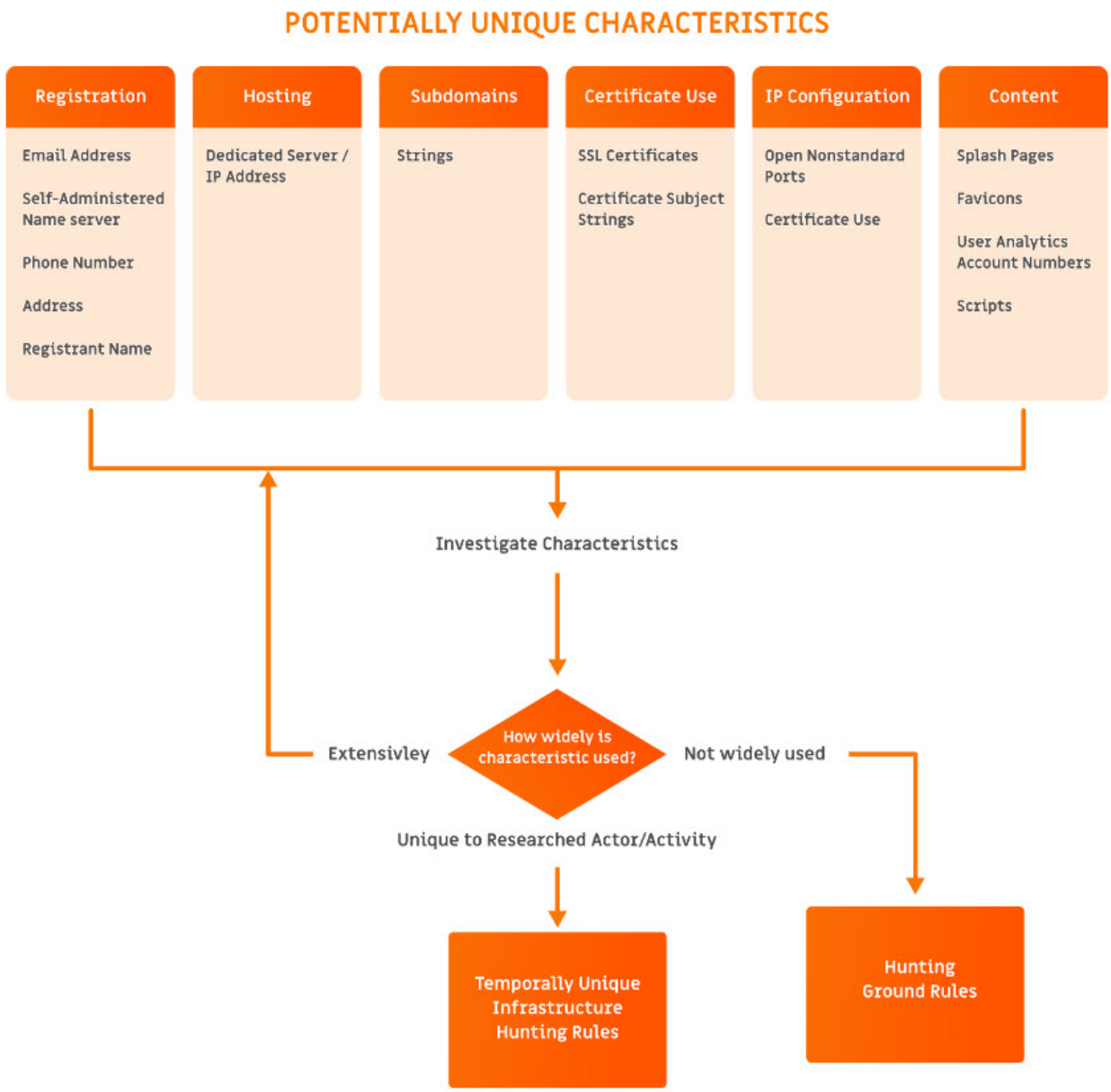
We would be remiss if we didn't at least mention that there is also a symbiotic relationship between the output from our infrastructure research and hunting efforts and the capabilities node of the diamond model. Malware analysis efforts can identify infrastructure in behavioral information that feeds infrastructure research for a given actor/activity. Conversely, output from our infrastructure hunting efforts can be compared against sandbox reports, used as fodder for malware hunting by way of YARA rules searching for the output, or scanning of identified infrastructure for hosted malicious files.

Infrastructure Research

In this section we'll describe, at a higher level, our workflows for infrastructure research and hunting when investigating infrastructure associated with an actor or activity.

Singular Characteristics

When starting an investigation into a given domain or set of domains, we generally follow the below workflow, and start by looking for single, unique characteristics. Please note that the below list is not exhaustive but is generally those that we've encountered regularly. (Side note: Determining whether a characteristic is "unique" generally requires that we perform additional research in one of the aforementioned tools to see how widely used that given characteristic is. Whenever there is uncertainty as to whether a given characteristic is unique, that should be considered in the analysis attributing a new domain to an actor/activity.)



Unique Characteristic Research Workflow

This really represents the base use case for infrastructure research — think back to using one adversary's registrant email address to find the other domains they created. But as WHOIS privacy protection, GDPR masking, and anonymous identity resellers (eg. Njalla) have become more prevalent over the last 2+ years, unique registration pivots have become harder to come by. The other dimensions as unique pivot points have therefore become more important and pertinent to investigations.

There is also an iterative component to this research as well — additional infrastructure identified from the characteristics of one dimension, may lead to new unique characteristics in the other dimensions.

The findings from this initial workflow focused on singular, unique characteristics can be immediately operationalized for infrastructure hunting, where new domains are monitored for the given characteristics and associated with the related actor/activity. However, there are important considerations longer term, specifically for IP addresses, subdomain and certificate strings, or content. For these dimensions, while the given characteristic may be unique now, that could change in the future as IP addresses are reallocated or strings are co-opted for use by other actors. All that is to say that it's important to know where unique characteristics today might not be unique tomorrow.

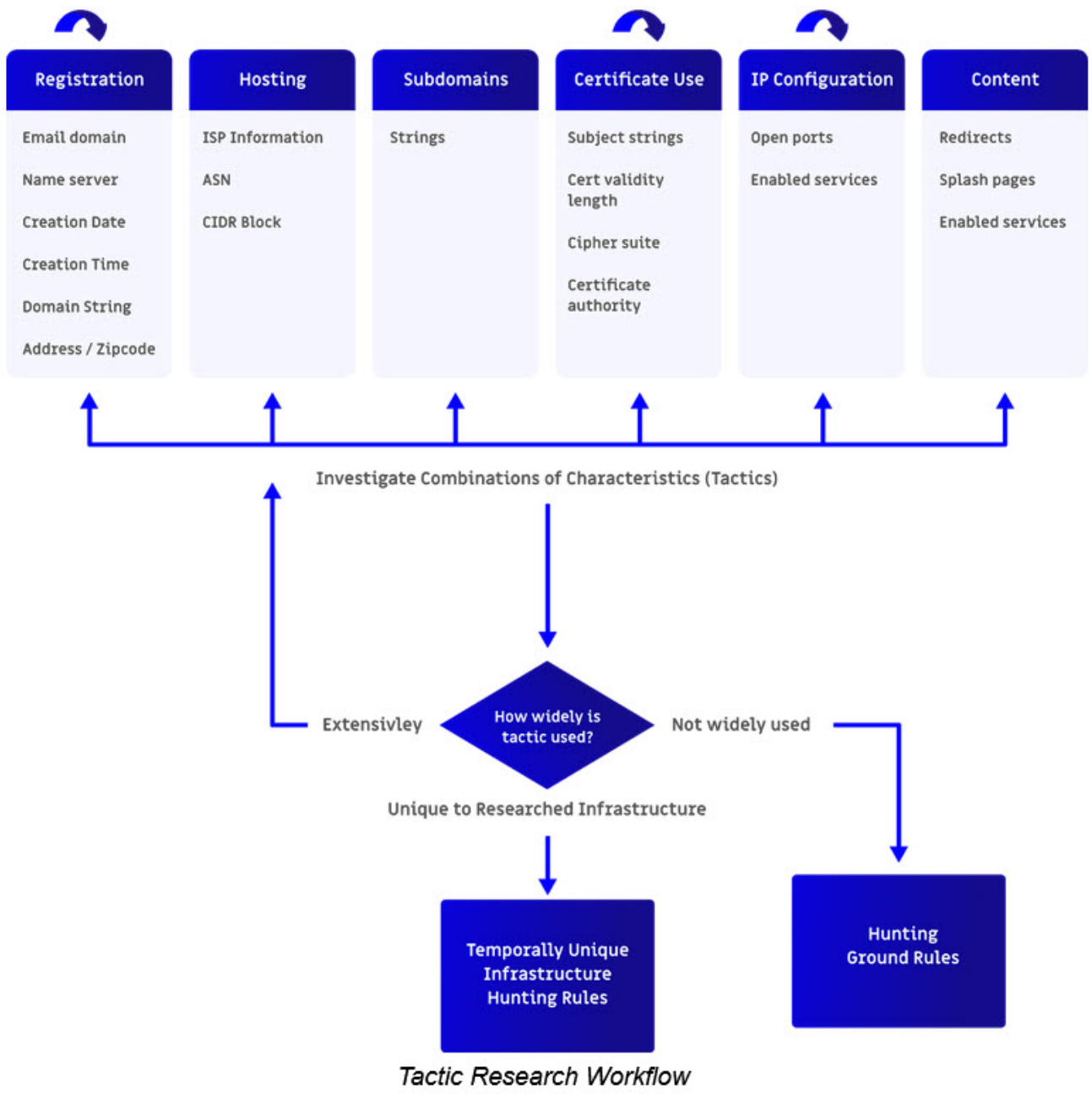
For notable singular characteristics that are not unique, but not widely used, such as the use of a small CIDR block to host domains, hunting can be geared around those characteristics to generate hunting ground against which additional analysis will yield the relevant infrastructure. Or those characteristics can be taken in conjunction with others, as described in *In the Land of Registration Protection, Characteristic Confluence Reigns.*

Examples of previous research based on unique characteristics:

- Dedicated servers, Subdomain strings: Kimsuky Phishing Operations Putting in Work (9/28/20)
- SSL certificate, Dedicated servers: Lights, Camera, Actionable Intelligence (3/18/19)
- Certificate strings: A Song of Intel and Fancy (3/16/18)
- Email address: Duping Doping Domains (1/11/18)
- SSL certificate, Splash pages: Track to the Future (9/21/17)
- Dedicated servers, Email address: Let's Get Fancy (10/18/16)

In the Land of Registration Protection, Characteristic Confluence Reigns

After exhausting our research into individual characteristics we'll then move to investigate combinations of non-unique characteristics. It is in this workflow that we've found the most investigative success since the advent of privacy and GDPR protection. The idea with this workflow is to identify as many different combinations of two or more non-unique infrastructure characteristics within or between the dimensions that, when taken in conjunction, potentially identify unique tactics or tactics that are not widely used.



Tactic Research Workflow

The characteristics shown above are those that we tend to review in conjunction with other characteristics to identify tactics within and among those dimensions. It should be noted explicitly that for the Registration, Certificate Use, and IP Configuration dimensions, tactics can be identified from entirely within that dimension also. As an example, an adversary may consistently register their domains using the same email domain and boutique name servers.

During this workflow, we are constantly checking and making note of how widely the given confluence of characteristics — or tactic — is present in infrastructure beyond that which we're investigating. Ultimately, like with the previous workflow, we're looking to identify tactics

that are either **1) unique to the infrastructure we're investigating** or **2) not widely used beyond the infrastructure we're investigating.** Keep those two in mind as we explore considerations in the *Infrastructure Hunting* section.

It's important to note that this research workflow is much less structured and instead is assisted by the researcher's knowledge of commonalities in both the infrastructure dimensions and actor/activity. As an example, knowing that specific ASNs are widely used in conjunction with domains registered through a specific registrar can help narrow down the characteristics of the given infrastructure that actually merit focus. Further, if a threat group is known to have used boutique email domains and reused resellers for procurement, those are characteristics this workflow can focus on. The more familiarity the researcher has with the given dimensions and researched actor/activity in general, the quicker they'll be able to spot the characteristics that merit consideration as part of a tactic.

Examples of previous research based on tactics:

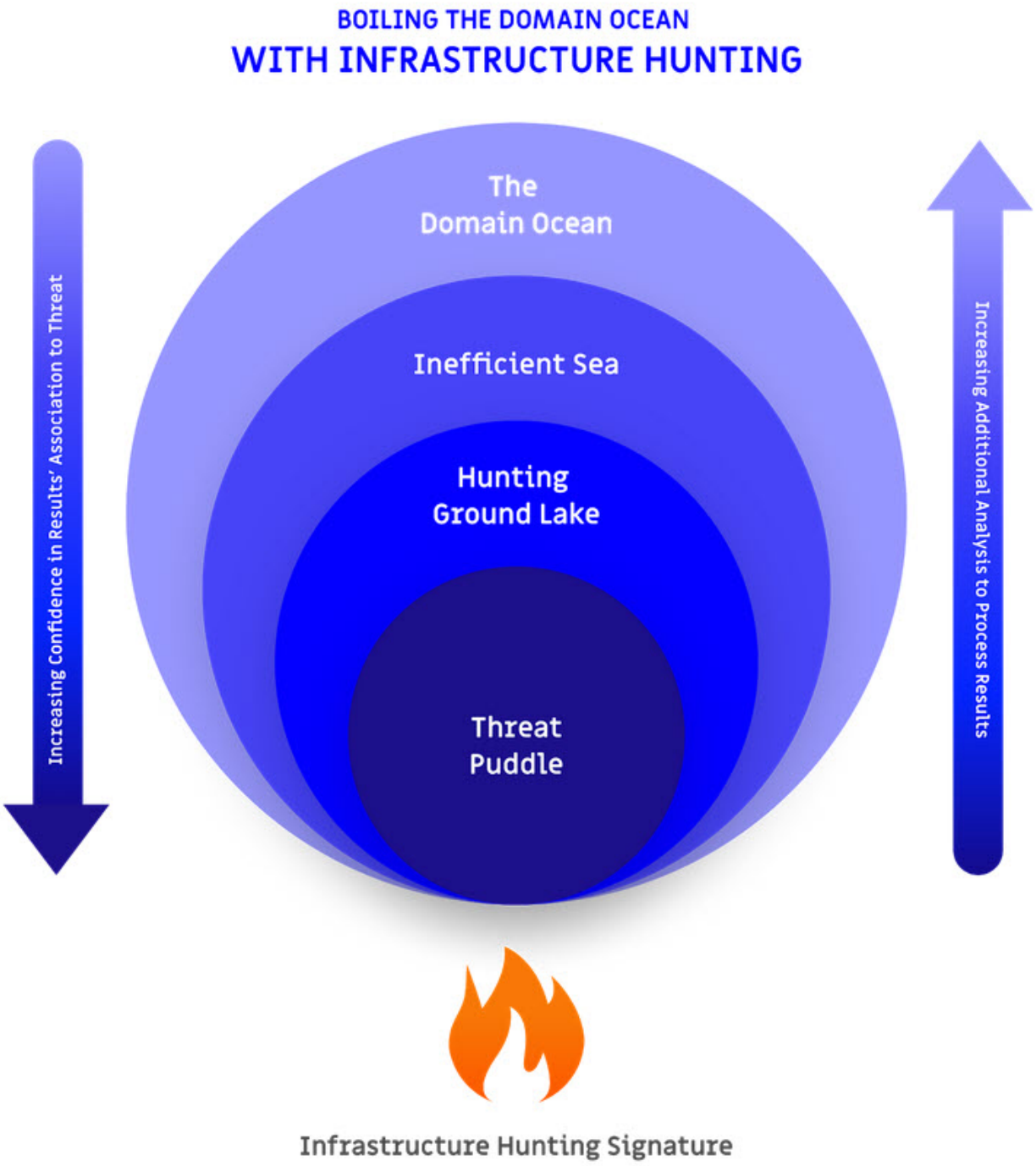
- Domain Strings, Name server, SSL Certificates, CIDR Block: 20201019A: Additional Ryuk Infrastructure (10/19/20)
- Creation Timestamp and Name server: Building Out ProtonMail Spoofed Infrastructure (7/26/19)
- Creation Date and Name server: 20200529A: Network of Probable Sandworm Infrastructure (5/29/20)
- Domain Strings, Email domain, Name server, SSL Certificates, Content: Domains Spoofing Ukrainian Gas and Media Companies (12/16/19)

Infrastructure Hunting

Following our research efforts, we can operationalize our findings by way of infrastructure hunting. We typically use the aforementioned tools to create and run the signatures / rules that are based on our research findings. When possible, we automate the execution and output ingestion of these rules in ThreatConnect to alert on new results, memorialize them, and kick off additional analysis. Check out a recent webinar we did with DomainTools for a specific example. This section captures infrastructure hunting considerations at a high level.

Effective Signatures

If we consider the body of domains that exist, with hundreds of thousands registered every day, it is a metaphorical ocean. With infrastructure hunting, our goal is to build a strong enough fire with the research-derived rules to boil that ocean down to a smaller body of water with domains that are specific to our threats or against which minimal additional analysis can identify their domains. Consider the below graphic to illustrate this concept behind infrastructure hunting and striving to generate effective rules.



Illustrated Concept of Infrastructure Hunting

Thinking back to the *Infrastructure Research* section, we mentioned that we were looking to identify characteristics and tactics that were **1) unique to the infrastructure we're investigating** or **2) not widely used beyond the infrastructure we're investigating.** The reason being is that either could be used to craft effective infrastructure hunting rules. Sure, the best case scenario is having signatures that exploit threat-specific characteristics and boil *The Domain Ocean* down to a *Threat Puddle* where we have high confidence that the resulting domains are associated with the threat we're concerned about.

However, rules that result in the *Hunting Ground Lake* and a small number of domains that both are and aren't associated with our threats still are valuable to analysts looking to proactively identify relevant infrastructure. With these results, we initially have less confidence in their association to the threat and have to conduct additional analysis against the output infrastructure to determine whether we can make the analytic judgement that an output domain is associated with the hunted threat.

How much tolerance a researcher or organization has in conducting that additional analysis against the hunting ground domains will vary. To that end, how specific those hunting rules resulting in this output need to be will vary accordingly. For example, one analyst hunting for a given threat might be able to tolerate conducting additional analysis against 50 domains a day to better identify their threat's specific domains, whereas another may be willing to tolerate 100. To that end, the size of the *Hunting Ground Lake* will vary for both and the latter analyst can craft less specific signatures to cast a wider net for their threat's domains.

What we want to avoid is rules that land us in *Inefficient Sea* (inefficiency, get it?) where there is little-to-no confidence that the excessive output domains are associated with our threats, and against which additional analysis would be too cumbersome to ultimately identify relevant domains.

Additional Analysis and "Suspicious" Domains

Further analysis against domains from hunting ground rules seeks to determine additional consistencies with an actor's previous operations. Processing the output could involve manual review of the domains for things that you couldn't include in the signature (think back to the other dimensions mentioned above), use of additional tools to identify other traces of activity or tactics employed, or building out an understanding of that domain's associated infrastructure for other actor consistencies. How much, if anything, you can identify while further analyzing infrastructure in the hunting ground is ultimately going to impact the confidence with which you can assess that an output domain is associated with an actor.

Despite our best efforts to enrich and further analyze domains from our hunting grounds, in many cases we may not find any additional context to help us determine whether an output domain is associated with a given threat. In these instances, we tend to label this infrastructure as "suspicious" given that it has non-unique consistencies with our threat's identified infrastructure tactics. We can still take defensive measures against these suspicious domains, and can develop longer term analytic processes (monitor DNS changes, regular content scans, YARA rules to hunt for files with relevant behavioral information) to monitor for new context and update our understanding of the infrastructure.

Multiple Signatures for Wider Coverage

During the course of your research, you may generate multiple rules exploiting various characteristics or tactics identified in a set of adversary/activity infrastructure. This is perfect and a great way to ensure that you'll still identify a threat's domains if they reuse one tactic but not another. We would caution against creating one signature to rule them all and exploit the various identified tactics as it becomes more difficult to understand what specific tactic resulted in an output domain. What can be useful is grouping related tactics (those exploiting the same characteristics in the same dimensions) together.

For example, if you identify that your threat has used Protonmail and Tutanota email addresses to register domains hosted in two ASNs, having a single rule to capture domains registered using either one of those two email domains AND either one of those two ASNs mitigates the need to create four separate rules for the related tactics.

Proactive Infrastructure Hunting and Defense Considerations

Infrastructure research and hunting efforts ideally move organizations toward a proactive stance with respect to their identified adversaries where defensive action is taken against those adversaries' infrastructure before it is operationalized against the organization. However, these efforts are not without additional considerations.

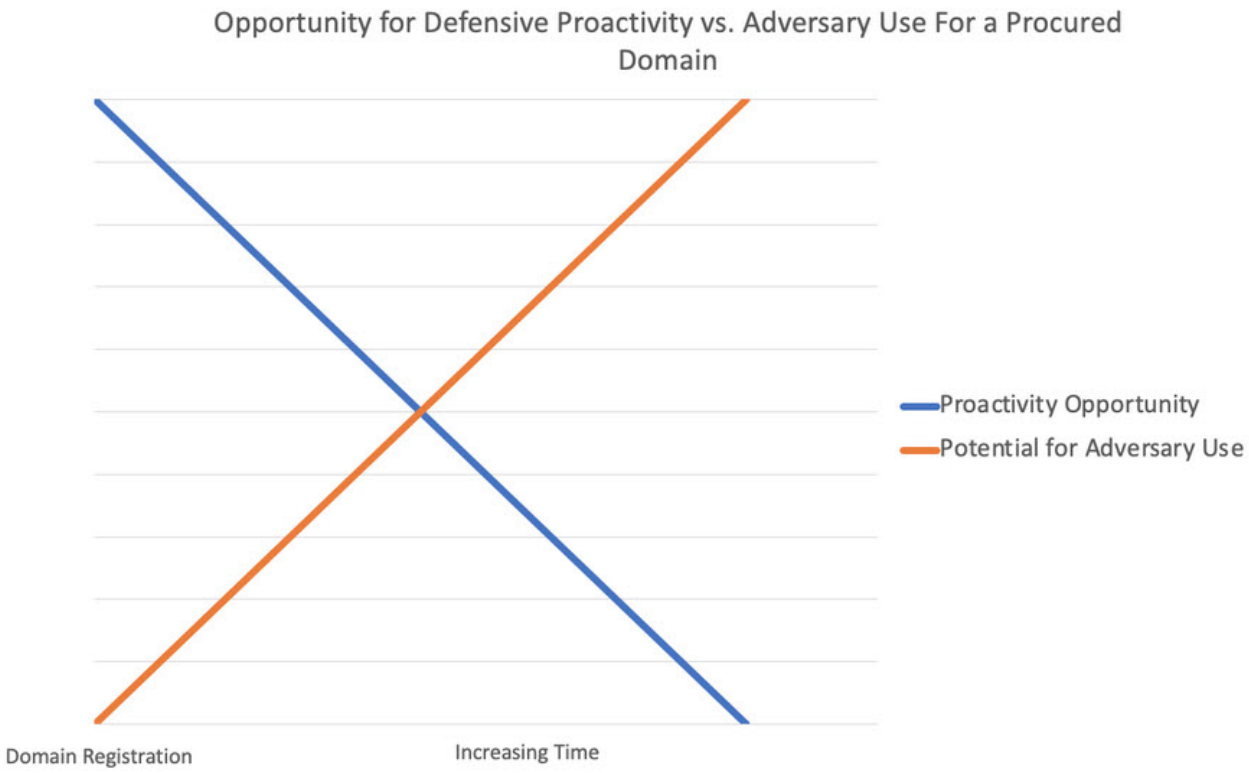
Defensive Action vs. Research

It's important to note that defensive action considerations for output from infrastructure hunting may differ in scope and tolerance from those heretofore described for research and analysis. For example, an organization may want to automatically block or monitor for every domain from a hunting rule that lands you in *Hunting Ground Lake*. The idea being that even those output domains from the given rule that aren't associated with the hunted threat still have no legitimate utility for a business, so taking proactive defensive measures against all would mitigate the time and work going into additional research against those domains to identify the threat-specific results.

As an example, if you identify that one of your threats regularly registers domains through a boutique reseller (as reflected in the name server) that only creates a couple dozen domains a day, automating defensive action against all those domains could be a viable, proactive measure, even if they aren't all related to the given threat.

Sacrificing Context for Proactivity

A notable aspect of leveraging infrastructure hunting output is that we generally have to sacrifice context for proactivity. However, we'd argue that is a sacrifice most organizations should be willing to make, especially for their most pertinent threats. Taking a "wait and see" approach to garner more information on the domains that your specific infrastructure hunting signatures yield invariably subjects an organization to unnecessary risk from those domains. Consider the below graph showing the inverse relationship between proactivity and adversary operationalization.



When a domain is initially registered, that is when an organization has the best opportunity to address that domain before the actor has a chance to use it against their organization. Over time, the chances that the adversary will have used the domain increase while the opportunity to proactively address it decrease accordingly. The better you can decrease the delta between when a potential adversary domain is registered and your defensive action is taken, the better, even if it's at the detriment of context.

Thanks From the Future

Another important consideration related to proactivity is that the new infrastructure that you discover from hunting rules and build out now might not be relevant until some time in the future. Said another way, a lack of actual current activity associated with the domains you researched doesn't mean that your research was done in vain. A suspicious domain today could be the command and control (C2) domain used in conjunction with your adversary's new malware tomorrow. When that's the case, having already built out an understanding of the suspicious domain and other associated infrastructure associated with it enables more thorough defensive actions from your future self.

Potential Pitfalls

Infrastructure research and hunting processes generally are not infallible and it's important to understand the common pitfalls to avoid during the course of your investigations.

Temporal Volatility of Hunting Signatures

As noted above, there is a temporal aspect to the uniqueness of many characteristics or tactics identified during the course of infrastructure research. Over time, by chance or intention, more malicious and benign actors may use characteristics and tactics that were once specific to the threats that you researched. Email addresses can be dropped and reused by others in the future; other individuals may coincidentally register domains that mimic a threat's registration consistencies. To that end, it's important to understand how the uniqueness of those characteristics degrade over time, how easy they are to mimic, and how those concerns ultimately affect confidence levels in attributing domains to the hunted threats.

Susceptibility to Co-opting / False Flags

There are other concerns regarding the volatility of infrastructure hunting rules, notably that other adversaries can co-opt known tactics to give the perception that the activity they are conducting is associated with another threat. One recent example of this was identified in an August 2020 Kaspersky report on DeathStalker where the actor had used an SSL certificate subject string consistent with previously identified Fancy Bear infrastructure. In addition to the aforementioned notes regarding volatility affecting confidence levels, this is an aspect of infrastructure hunting that multiple rules, thorough knowledge of the adversary's previous infrastructure, and capability node centric analysis can help avoid. As an example, if a domain hits on a hunting rule that has previously been specific to your adversary, but doesn't carry any of the other characteristics previously seen with that adversary or the output domains from that rule, then that should ultimately affect your confidence level in and language associating the domain with the adversary.

Erroneous / Excessive Pivoting

Finally, the biggest pitfall that someone can run into when getting into infrastructure research and hunting is erroneous pivoting — using a non-unique characteristic as though it is a unique characteristic. Most often, this manifests in pivots based on hosting IPs where the researcher mistakenly believes that all domains at a given IP are associated with the domain they are investigating. Sometimes these IPs can be parking lots, sinkholes, or non-dedicated infrastructure that isn't widely used but seems exceedingly suspicious and therefore all related. The same can happen for other characteristics as well, so it is important to thoroughly consider all the characteristics upon which you're building out adversary infrastructure. Always investigate the uniqueness of those characteristics and ensure that any uncertainty is reflected in the confidence levels in your analysis.

Conclusion

This blog post was intended to provide a higher level overview of how we go about infrastructure research and hunting, specifically as it relates to procured domains. There are so many investigative avenues in those dimensions that we mentioned (and almost certainly

others) that can help you build out an understanding of your adversaries' infrastructure and proactively hunt for and defend against them.

If you're looking for more specifics on how we've employed these processes before, consider checking out some of the blogs previously referenced or logging into ThreatConnect to see the recent research we've shared.