A Cyber Threat Intelligence Self-Study Plan: Part 2

medium.com/katies-five-cents/a-cyber-threat-intelligence-self-study-plan-part-2-d04b7a529d36

Katie Nickels August 23, 2022

Katie Nickels

Aug 22

.

11 min read

It's been a while, but here it is — part 2 of my cyber threat intelligence (CTI) self-study plan. This is much later than I intended, but life happens. It has been wonderful to hear feedback on how part 1 helped many people learn more about CTI, and I truly appreciate everyone who reached out to say that this helped them. You all are the reason I created part 2, and I hope it's just as helpful. For background on this plan and how I recommend using it, please read my first post. This post covers four topics:

- OSINT and Open Sources
- Pivoting
- Clustering, Naming, and Creating Groups
- Attribution

Let's dive right in!

OSINT and Open Sources

Open source intelligence (OSINT) is a separate but complementary field to CTI. As you learned in part 1, intelligence is analyzed information to support a decision, so it's important to remember that definition still applies. I think of OSINT as being intelligence produced using open sources, though it's commonly used to refer to open source data or information. It's particularly important to critically evaluate open sources before using them to produce intelligence.

Read and watch

- Read selections from this Army publication on OSINT: —you can skim the whole thing, but I recommend reading just pages 21–23 (starting at Open-Source Reliability heading), page 41 (Information Reliability and Credibility heading), and page 43 (section 4–13) (hat tip to Casey Brooks for this reference)
- Read about how to differentiate reports from vendors as threat intelligence or marketing using the ADEPT model.
- Skim , focusing on the OSINT definition and the other 5 “INTs”.
- Watch this from Sherman Chu on Current Intelligence.
- Skim this page that lists many available open sources: . Think about the major categories of sources and where you think this data comes from.

Things to do

- If you haven’t already, create a Twitter account and follow some CTI people (if you follow me @likethecoins, Twitter will automagically start recommending other accounts to follow). I also recommend creating your own RSS feed as recommended in , but if you don’t want to, here is that you can import into any RSS feed tool of your choice. (Feedly is a good starting point!) Please note this is a recommended starting point, and I’m certain I forgot some excellent blogs, so I appreciate your understanding if yours isn’t included.
- Find a blog post or report from a cybersecurity vendor via your Twitter or RSS feed. Apply Sergio’s ADEPT model and make your own assessment on whether it falls more under intelligence and marketing.

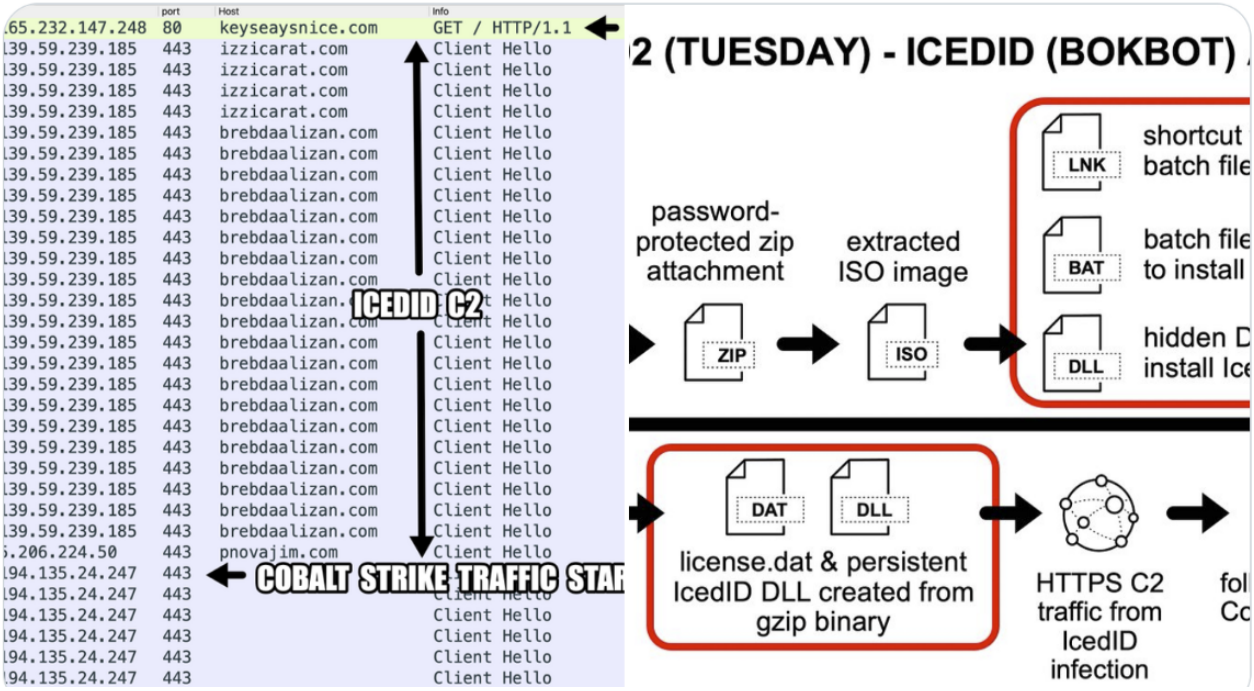
4-13. When evaluating sources of information to determine reliability and credibility consider—

- **Identity.** *Who* produced the information (for example a student, teacher, political organization, or reporter)?
- **Authority.** *How* much does the source know about the information?
- **Motive.** *Why* was the information published?
- **Access.** Did the source have direct access to the event or information?
- **Timeliness.** *What* is the date of the information?
- **Internal and external consistency.** Does the information contradict governmental policies among local citizens?

- Find a different blog post or report and evaluate its reliability and credibility based on section 4–13 of the Army PDF above — Identity, Authority, Motive, Access, Timeliness, Internal and External Consistency (note that you will likely apply the Internal and External Consistency criteria differently than the Army publication. Think of internal consistency as whether the piece contradicts itself. Think of external consistency as whether the piece has contradictory information to other external sources.).
- Find a Twitter account and apply the same criteria above to an account and tweet. Here’s an example for the following tweet:

Brad
@malware_traffic

2022-08-02 (Tuesday) - Again with #CobaltStrike on 194.135.24[.]243:443 after an #IcedID infection. First reported by @drb_ra on Tuesday 2022-07-26. I saw it 2 days later and tweeted about it on Thursday 2022-07-28. Happened again yesterday. Happened again today.



> **Identity:** With a little searching, I can determine this account is owned by Brad Duncan, who is a researcher at Palo Alto's Unit 42.
>
> **Authority:** I have read blog posts from Unit 42 in the past that have been well written, so that makes me more likely to trust this account if I've never seen it before. I look through Brad's previous tweets and see he has a long history of tweeting about various malware and threat actors. He has over 65K followers, which is a lot, but that's not the only point I'm using to evaluate his authority.
>
> **Motive:** I see Brad has a whole website full of packet capture and malware samples: https://www.malware-traffic-analysis.net/. Combined with all his previous tweets and his replies to other researchers, I assess he is someone who shares information because he wants to help others track threats. I should remember he also works for Unit 42, but the fact his website is separate suggests he isn't motivated by marketing for his employer.
>
> **Access:** The fact that Brad showed screenshots of the packet capture suggests he did have direct access on this network traffic.
>
> **Timeliness:** This is very timeline, he tweeted this information the same day the traffic occurred.
>
> **Internal and external consistency:** Brad spelled everything correctly in his tweet and diagrams, and nothing in the information contradicts itself. Additionally, Brad cited another researcher, @drb_ra, who reported the same findings.

Questions to think about

- Based on the DNI webpage, how does the U.S. government think about OSINT? How does that differ from how the private sector should think about and practice OSINT?
- What are the risks and drawbacks of using open sources in cyber threat intelligence analysis? How can you mitigate those in order to use open sources in an appropriate way?
- Who produces what is available in different open sources (for example, the sources listed in Awesome Threat Intelligence)? What limitations does that result in?
- What CTI requirements (see part 1 for more on requirements) do you think can be fulfilled with open sources? How do you think open sources can be best used to produce CTI?

Pivoting

I define pivoting as the act of identifying a new piece of information based on a connection to a previously-known piece of information. CTI analysts apply pivoting in many ways, such as looking for connections between domain names and IP addresses that have hosted them.

Sometimes, these connections can reveal new relationships between groups or malware that analysts weren't previously aware of. Pivoting can be challenging to learn because it requires both an understanding of the data source an analyst wants to pivot through as well as methodology for doing so. I've included resources that provide explanations for common data sources used for pivoting, but note that you may need to do more research to deeply understand these sources.

Read and watch

- Watch a short clip of a I did (18:37–21:02).
- Read Joe Slowik's on "Analyzing Network Infrastructure as Composite Objects", which presents three pivotable artifact types: domain names, IP addresses, and TLS certificates.
- Read Kelsey LaBelle's blog post series on , , and , which are all commonly-used sources for pivoting.
- Watch Mark Parsons' on TLS certificates and/or read his .
- Read Joe Slowik's on "Formulating a Robust Pivoting Methodology" and/or watch his on "Pivoting from Art to Science."

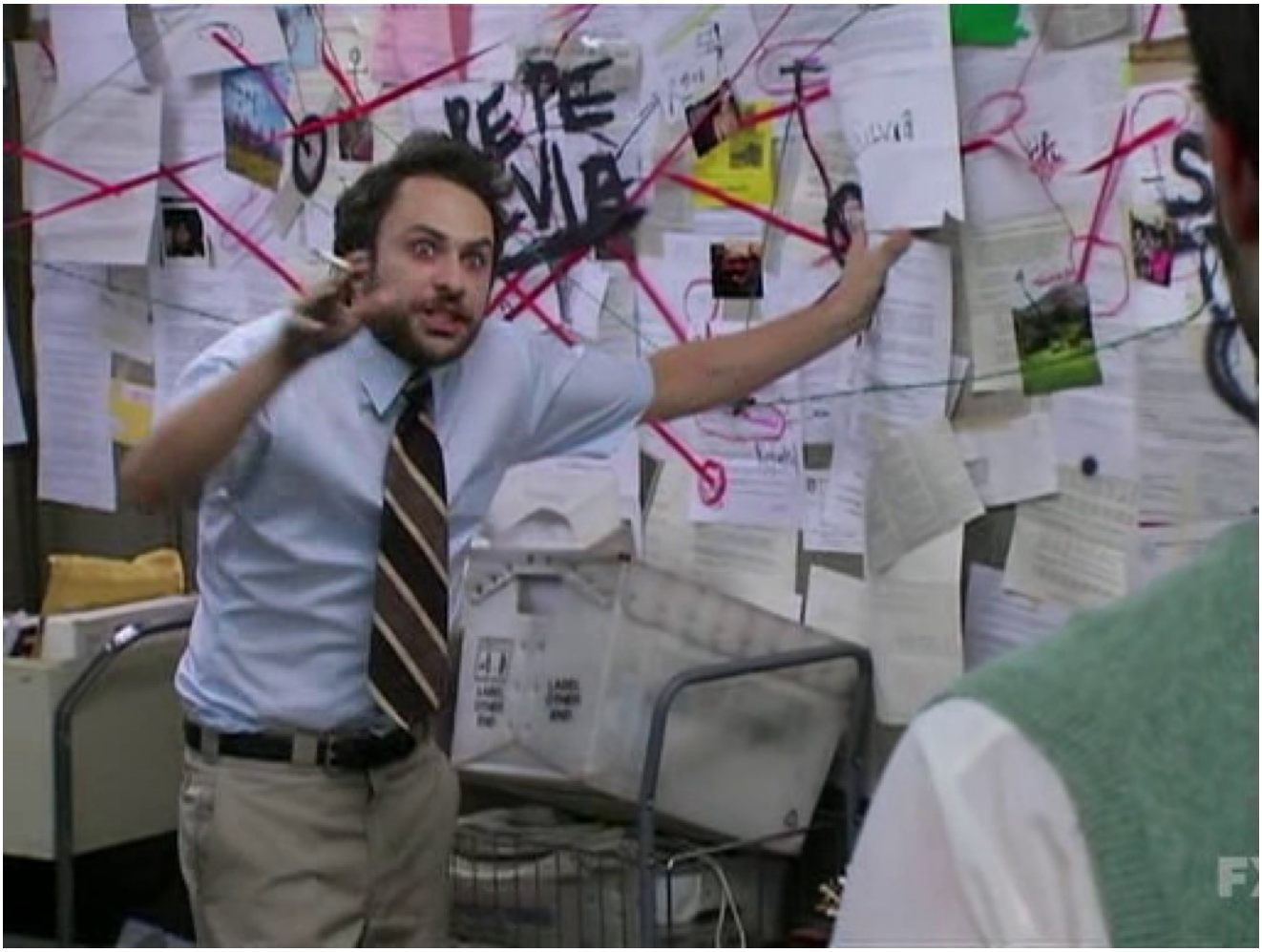
Things to do

- Try a TLS certificate pivot by…1. Identifying the default SHA256 hash value shipped on Cobalt Strike servers (see), 2. Searching for that hash value in Censys (), and3. Identifying IP addresses that result. What do those IP addresses potentially represent? What could you do to try to validate if that pivot you performed was meaningful? Go do it!
- Find a recent blog post on a threat group or malware family that contains some pivotable artifact (look at the RSS feed you hopefully set up based on). Try searching for those artifacts in free search sources and see if you can reproduce the research from the blog post about any pivots and/or find any new artifacts from pivoting.

> Here are some free search sources you might find useful for this: Censys (https://search.censys.io/), Shodan (https://www.shodan.io/), DomainTools WhoIs (https://whois.domaintools.com/), VirusTotal (https://www.virustotal.com/), URLhaus (https://urlhaus.abuse.ch/), and Malware Domain List (https://www.malwaredomainlist.com/). There are many others listed on Awesome Threat Intelligence that you can try as well.
>
> If you need an example to get you started, try looking up hash values and IPs from this ESET blog post in VirusTotal.

Questions to think about

- What types of artifacts and data sources can we use for pivoting? What artifacts do you see most analysts use? What are the pros and cons of each type of artifact/data source? Can you think of data sources that could be useful for pivoting that weren't mentioned in the above resources?
- How do we know if pivots are meaningful or not? What level of specificity or uniqueness makes for a meaningful pivot?
- What is the point of pivoting? What do analysts gain from it?
- What are potential consequences of making analytic conclusions based solely on pivots?
- What happens as you continue pivoting off of newly-found information? How does each "hop" from your initial data point affect the relationship of the newly-found information? What are the risks and limitations of pivoting? (Hint: see below. Also see the Attribution section and Sarah Jones' presentation on attribution mistakes.)



is a cautionary tale of what happens when you pivot too far

Clustering, Creating, and Naming Groups

There are many ways to create clusters and threat groups, and this leads to confusion in the community as well as a lot of different names. While this can be frustrating for many analysts, hopefully you can reduce this frustration by gaining a better understanding of how CTI analysts create and name these groups.

Read and watch

- Watch part of I did that discusses group naming with the Diamond Model using the Rule of 2. (watch from 37:12–41:43)
- Watch I did on a deep-dive into group naming.
- Read Florian Roth's on group names.
- Read these blog posts on how various teams name groups: (and), Microsoft (first paragraph and take a look at the nation state group names on page 51), and .
- Read on UNC2452 and APT29.

Things to do

- Choose a group from the that has multiple associated names. Read the reference(s) for that associated name overlapping with the "main" group page name. Note what details, if any, explain the reason for the overlap.
- Using the methodology of your choice (a spreadsheet might be a good option), organize data points from the following information on various IcedID campaigns: , , , and . Considering using the Diamond Model and Rule of 2 to do this (see above link for background). Based on the data you have, decide if you think you should create a group (since IcedID is malware, you could create a group for the delivery of it). If you create a group, create a definition for that group — in other words, if you saw a new IcedID intrusion, what criteria would it need to meet in order to be part of the new group you created? Think about what challenges you had in doing this and what criteria is "close enough" to create a group or not.
- Search Twitter to find a recent debate about group names in which analysts disagree (this is a debate that comes up a lot in the CTI community). Think about each party's perspective and note reasons why they likely hold that view, even if you don't agree. (Hint: search Twitter for APT31, APT10, or TA410.)

Questions to think about

- Do you think Rosetta Stones are useful? (A is a resource that "translates" different group names, for example, .) What are the limitations of Rosetta Stones for tracking group names?
- Do you think the CTI community can agree on a consistent names and definition for groups? Why or why not? How could this happen if analysts wanted to try?

- What are the risks of using a group name that another team created? (Here's .) Do you think that risk is worth it to not add another group name (and potentially more confusion)?
- What are some common attributes/observables/data points that analysts use when identifying overlaps to create groups? (The Spooky RYUK-y webinar is a great place to look for this.) Which of these attributes do you think are most and least useful for identifying meaningful overlaps? (for example, if you saw PowerShell used in two different incidents, would you create a group based solely on that technique use?)
- What do you like and not like about methodologies used by different teams to cluster? What are the benefits and drawbacks of using an uncategorized clustering methodology for initial analysis? (e.g.)

Attribution

Attribution is a popular topic in CTI. While it is the primary focus for some analysts, it may not be for others. It's important to understand some of the nuances of attribution and how it's done, as well as how it can be useful (and when it might not be). Attribution is a topic that intersects cybersecurity and international relations, as an attribution assessment can have a significant impact on the relationships between countries. There are many articles on attribution from academia, and I've just chosen a sampling below.

Read and watch

- Read I wrote on types of attribution.
- Read from Steve Miller on Detectrum.
- Read from Robert M. Lee.
- Watch Sarah Jones' on "A Brief History of Attribution Mistakes." Tying this back to an earlier section, think about the role pivoting played in these mistakes.
- Read from Jason Healey that includes the ever-useful Spectrum of State Responsibility.
- Skim from Thomas Rid and Ben Buchanan — if you're short on time, just look at the detailed Q Model on page 31.
- Read , "The Purposes of U.S. Government Public Cyber Attribution," from Jon Bateman.

Things to do

- Choose an indictment from . Look at the indictment itself (PDF format) and write down the key evidence/information used to associate activity to an individual or group.

- Examine the IcedID campaigns from the section on clustering. Write down what additional information you think you would need in order to try to attribute these campaigns to 1. A country, and 2. A person. You could use the information types from the indictment you analyzed above, or other sources you can think of.

Questions to think about

- How is attribution commonly thought about in the cybersecurity community? What could be different ways to look at or define attribution?
- Do you think attribution matters? Under what circumstances would certain types of attribution matter or not matter? What requirements would result in CTI analysts deciding to do different types of attribution?
- What are the differences between the public and private sectors in 1. The process of conducting attribution (available data sources), and 2. Requirements to conduct “true” attribution (to a country, military unit, or person)?
- What are the risks of performing attribution, especially attribution to a country, military unit, or person?
- What are some of the challenges of performing attribution?

That’s all I have for now! I make no promises on Part 3, but the more I hear that this helps people, the more likely it is that will happen. :)