CTI is Better Served with Context: Getting better value from IOCs

klrgrz.medium.com/cti-is-better-served-with-context-getting-better-value-from-iocs-496343741f80

Andy Piazza December 3, 2020

Andy Piazza

Jun 14, 2020

.

9 min read

Here’s my GIANT head, drawn by this amazing artist:

Whether you are disseminating threat indicators internally to other teams or participating in information sharing programs within the community, context is a critical component of actionable intelligence. When analysts say, “Indicators aren’t Intelligence”, they are often referring to the contextless sharing of Observables that is too common within the Cyber Threat Intelligence (CTI) community. I believe that indicators can provide serious intelligence value, but it is up to the source analysts to provide that worth in the form of context.

Administrative note: In this article, I will capitalize key words to indicate field names and to stress the difference in terminology. Sorry tech-writer friends, you can blame the Army for giving me the Bad Habit.

Observable or Indicator of Compromise?

Before we go too far in the discussion of information sharing standards, I need to identify terminology because a lot of these words are used interchangeably in everyday discussions, which can lead to confusion. So let’s talk about the difference between Observables and Indicators of Compromise (IOC). In my view of the CTI world, Observables are the Atomic and Computed indicator types discussed in Lockheed Martin’s Cyber Kill Chain Paper. These are values found in system logs, network traffic, or computational outputs. Without including human-assessed context, these are just Observables. Examples include hashes, IP/Domain/URL, Registry Keys, etc. This also includes the metadata around these Observables, including timestamps, ports, protocols, etc.

Here are Lockheed’s definitions, which are found in section 3.1 of their research paper Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chain:

- “ — Atomic indicators are those which cannot be broken down into smaller parts and retain their meaning in the context of an intrusion. Typical examples here are IP addresses, email addresses, and vulnerability identifiers.
- — Computed indicators are those which are derived from data involved in an incident. Common computed indicators include hash values and regular expressions.
- — Behavioral indicators are collections of computed and atomic indicators, often subject to qualification by quantity and possibly combinatorial logic. An example would be a statement such as ”the intruder would initially used a backdoor which generated network traffic matching [regular expression] at the rate of [some frequency] to [some IP address], and then replace it with one matching the MD5 hash [value] once access was established.””

My opinion is that the community is better served by referring to Atomic and Computed indicators as Observables. To me, the Behavioral indicators are what I consider Indicators of Compromise (IOC), which are Observables with context.

Let's look at a traditional crime scene as an example. Fingerprints. For the sake of argument, fingerprints are a Computed indicator since you'll need to use some dusting powder and tape to get a good print. My fingerprints in your house are just an Observable. A police officer isn't going to lift my prints from your house and send them out to every other police station without including a note about them. That is because my prints do not indicate a compromise by themselves. What is the context behind how my fingerprints ended up in your house? Did you invite me inside and my fingerprints were found on a cold beverage can? Or were my fingerprints found on a rock that appears to have gone through your recently broken window? See how critical this context is to make an Observable into an Indicator of Compromise?

Well, the same concept applies to computer Observables. I get a feed of suspicious IPs into my TIP and I push them into my SIEM to generate notable alerts to go look at. When I see that an IP from a threat feed matches an IP in a proxy event, I look at the IP in the threat feed and find it doesn't have any context about how it was observed and reported in previous malicious/suspicious events. I don't have an indicator. I have an Observable. That is the equivalent of the police forwarding copies of evidence from cases to all other police stations, but not including the notes behind the evidence. It is extremely frustrating that we accept this — and we do it ourselves — within the CTI community.

So say it with me:

IOC = Observable + Context.

Within information sharing standards, this context is traditionally found in the Description field of an IOC. Hopefully, the context is enough to provide a basic understanding of what type of activity the IOC is related to and why it was shared. And no, I don't mean that "C2 IP" is good enough context. Threat Vendors, I'm looking at you! I shouldn't have to pivot to a threat report or blog to get context around an IOC. I also shouldn't have to copy/paste IOCs out of a PDF or manually type them from an image… but that's a separate rant.

> **PRO-TIP:** if your Threat Intelligence Platform (TIP) doesn't have a Description attribute for IOCs, you don't have a TIP. You have an Observable Aggregator.

Source:
When considering what to put in the Description field, there is an obvious balance between efficiency of processing and the effectiveness of context that is provided. I recommend a little Golden Rule of CTI: provide as much context as you wish other sharing partners would provide to you. Seriously, picture another analyst receiving your IOC through an information sharing program, they run those IOCs through their SIEM and get a match to an event and they have the Description you wrote. Did you give them enough information to determine if this is a False-Positive? Did you give them enough information to pivot into other event types to search for related IOCs? Or did you share another frustratingly contextless Observable?

> **Golden Rule of CTI: provide as much context as you wish other sharing partners would provide to you.**

A Word Please…

Before I go too far into laying out the minimum standards for IOC context. Let me take a pause to clarify three key points:

1 | When I'm talking about IOC context and the Description field, I specifically mean what gets included with an IOC in an indicator package, whether that is a CSV that accompanies a threat report or a threat feed. These IOCs likely have a lot of other great and relevant context in the source evidence (e.g. incident ticket or threat report). Source reports should maintain their standards of context, including screenshots of malware analysis, phishing email examples, etc. This is something that the community is doing well currently. Where I see it lacking though, is often times you get an amazing write-up and a table of IOCs without any direct link to where the IOCs fit in the kill chain. If you mention multiple files in a malware campaign, then list 20 different hashes, how is anyone supposed to know which part of the malware's execution those IOCs belong to? This is the perfect context for the Description field. Example: "This hash is the deobfuscated DLL of a Sneaky.Trojan campaign observed in June 2020." That is 100% better than a hash with the word "dropper" next to it in a table.

2 | IOCs need their own Description separate from the context of the source evidence because they are often carved out and sent to systems without any context at all. Adding context to the Description field in your sharing product enables other analysts to push those Descriptions with the IOCs into their SIEM. When traffic matches an IOC from the TIP, an analyst can see the appropriate context along with the alert. It drives me insane when I have an alert in a SIEM and I have to backtrack to the TIP first to find out that the source report's context isn't in the TIP and I have to go find a source report on the internet to find out why this IOC was in my TIP in the first place. That is craziness and it completely removes any value in having a TIP, threat feeds, and integration. This will occur for about a month before analysts stop doing the backtracking research and just begin marking all events as False-Positives.

3 | Getting everyone on board with the context levels I outline below is a pipedream. I get that nobody wants to do this kind of work. I know, I wrote Descriptions like this for four years to support an information-sharing product line. But I truly believe that this is the only way to get value out of information-sharing feeds. We are wasting time and money having analysts chase context in the middle of triaging potentially malicious events. We can do the work up front and save millions of wasted hours across the community.

Well… if you're still onboard with this article, let me show you what I would like to see in the Description field of more threat feeds. I'll start with the ideal state in the "Awesome Context" section and work down to the "Minimal Context".

Awesome Context

> *On 6/10/2020 "spoofed email[@]spoofedomain[.]com" sent "Invoice" themed emails that included the malicious URL "hxxp://totallynotlegit[.]com/probablymalware[.]exe" [MD5: 32 characters] (*T1566.002*), which is detected as Trojan.SuperMalware. When executed, "probablymalware[.]exe" [MD5: 32 characters] creates a Scheduled Task named "DoingBadStuff" (*T1053.005*) and begins C2 over port 443 to "hxxp://malwareC2[.]net" (*T1071.001*).*

> **PRO-TIP:** Putting your IOCs in "quotes" helps a lot with readability, especially when you include a full email subject or other multiple-word strings.

To me, this is an awesome level of context to provide for each IOC. If you're short on time, you can write this Description once, and copy/paste it into the Description field for the related IOCs. Let's be honest, if you're a CTI analyst, you are already a copy/paste pro. If you some extra time, you can include additional context for specific IOCs in the Description field for those IOCs. For example, in the Description for the Malware Hash IOC, you may add what Registry Keys are manipulated, or information about reverse engineering the payload. This isn't necessary to provide for the Email IOC, but is valuable context to add at the file level. Serious bonus points to all of the threat analysts that are adding MITRE ATT&CK technique references in their information sharing products. This is a great step towards a new high-water mark for CTI maturity. I see you and I appreciate you!

Great Context

> *On 6/10/2020 "spoofed email[@]spoofedomain[.]com" sent "Invoice" themed emails that included the malicious URL "hxxp://totallynotlegit[.]com/probablymalware[.]exe" [MD5: 32 characters], which is detected as Trojan.SuperMalware. When executed, "probablymalware[.]exe" [MD5: 32 characters] creates a Scheduled Task named "DoingBadStuff" and begins C2 over port 443 to "hxxp://malwareC2[.]net".*

Okay, this is really awesome context too. It is just missing the MITRE ATT&CK technique mapping. Two years ago, I would have said that this is the standard for awesome context. But after spending the last year or two mapping reports and IOCs to ATT&CK manually, I am loving that more and more CTI reports are including technique numbers. So this is great context instead of awesome context.

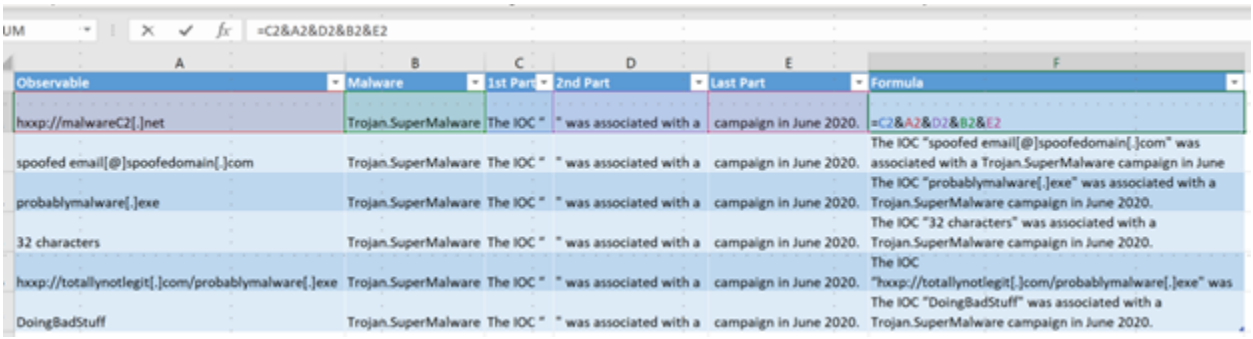
Minimal Context

> *"The IOC "hxxp://malwareC2[.]net" was associated with a Trojan.SuperMalware campaign in June 2020."*

This Description is arguably light on details. However, it is still a significant improvement over the context you get from most commercial threat feeds. Seriously, who decided that and Observable with a tag like “Malware” or “C2” was good enough? Do you know what happens if a TIP doesn’t have a Tag field? The association to the malware family is lost. We can do better than that for each other.

I get it. Sometimes you get a big list of IOCs to share and you don’t have the time (or don’t want to take the time…) to write a unique description for each IOC. In those cases, I recommend dumping the IOCs into a spreadsheet and using a simple formula to write a basic description automatically. Once you provide a basic template for the sentence structure and formula, you drag both the sentence structure and the formula down the spreadsheet to auto-populate a Description sentence. Here’s what that looks like:

- Column A: Observable (hxxp://malwareC2[.]net)
- Column B: Malware or Campaign Name (Trojan.SuperMalware)
- Column C: first part of a sentence (The IOC “)
- Column D: second part of a sentence. (“ was associated with a)
- Column E: Last part of a sentence. (campaign in June 2020.)
- Column F: =C2&A2&D2&B2&E2. (The IOC “hxxp://malwareC2[.]net” was associated with a Trojan.SuperMalware campaign in June 2020.)



UM =C2&A2&D2&B2&E2

	A	B	C	D	E	F
	Observable	Malware	1st Part	2nd Part	Last Part	Formula
	hxxp://malwareC2[.]net	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	=C2&A2&D2&B2&E2
	spoofed email[@]spoofedomain[.]com	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	The IOC "spoofed email[@]spoofedomain[.]com" was associated with a Trojan.SuperMalware campaign in June
	probablymalware[.]exe	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	The IOC "probablymalware[.]exe" was associated with a Trojan.SuperMalware campaign in June 2020.
	32 characters	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	The IOC "32 characters" was associated with a Trojan.SuperMalware campaign in June 2020.
	hxxp://totallynotlegit[.]com/probablymalware[.]exe	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	The IOC "hxxp://totallynotlegit[.]com/probablymalware[.]exe" was
	DoingBadStuff	Trojan.SuperMalware	The IOC "	" was associated with a	campaign in June 2020.	The IOC "DoingBadStuff" was associated with a Trojan.SuperMalware campaign in June 2020.

Excel formula in cell F2.

Final Thoughts

I encourage CTI analysts to consider this type of context when evaluating their information sharing standards, as well as during reviews of their collection sets. Hopefully, you have a Collection Management Framework and you’ve considered assessing your collection sources for value. I previously wrote about Threat Feed Assessments and that’s linked below. I recommend adding an assessment process that includes evaluating sources for contextual value. A source that provides a thorough Description saves you time and money during detection and response efforts. A source that provides a big noisy list of Observables without meaningful context is costing your organization money — even if the feed is free.

If this article was helpful or you completely disagree, either way please reach out on Twitter or LinkedIn to discuss. I love feedback and discussing concepts that lead to the maturity of the CTI community. Until then, happy hunting.

References:

Considerations for Leveraging Cyber Threat Feeds Effectively

My lessons learned and recommendations from multiple threat feed and Threat Intelligence Platform (TIP) assessments.

medium.com

https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/cyber/LM-White-Paper-Intel-Driven-Defense.pdf