Cyber Attribution Is More Art Than Science. This Researcher Has a Plan to Change That

therecord.media/cyber-attribution-is-more-art-than-science-this-researcher-has-a-plan-to-change-that/

February 19, 2021

Adam Janofsky
February 19, 2021

When Timo Steffens got his first job in threat intelligence more than a decade ago, with Germany's Federal Office for Information Security, or BSI, he had only a passing knowledge of advanced persistent threat groups and other hacking concepts that he's now an expert in.

His background was in artificial intelligence, and the agency, which is responsible for the German government's cybersecurity, was working on an early warning system to detect computer worms. Before his interview, he read through two books on cybersecurity, which sparked a passion for hunting down adversaries. Since then, he's written his own book on how to identify advanced persistent threat actors behind notorious cyber espionage campaigns.

The Record caught up with Steffens recently to talk about the latest in threat hunting, and why he thinks it's similar to disciplines like art history and criminology. He also discussed why the whole concept of APT groups—the term security professionals often use to categorize

nation state or state-sponsored hacking operations—is a simplification that the industry might have to move past. The conversation below has been lightly edited for space and clarity.

Timo Steffens

The Record: What's the benefit of attribution? What can be gained from knowing who attacked you?

Timo Steffens: When you look at the benefit, it depends who you are. When you're talking to security teams, they have different benefits than politicians and government agencies.

I think it's also a general psychological phenomenon that people just want to know who did something. I worked quite a few years as an incident handler, and when I came to sites and talked to victims, one of the first questions was always who was it and why did they do it—the technical means and how they got into the network were often secondary. And it was useful in that role to know who the actors might be and infer motivation and what they're after on the network… In the real world, you have a limited set of people and money, and you have to prioritize. If you know what some countries are strategically after, and your company is doing a joint venture in some region, attribution certainly influences your decisions on how to prioritize your security measures.

The last thing is around diplomacy and governments. Attribution is a prerequisite and necessary for anything diplomatic.

TR: Can you talk more about how attribution would influence decisions? If you found out an attack was carried out by a group like APT29, how would it be different from another group?

TS: When you're an incident handler and you talk to people and companies, the managers get really nervous that their production networks and [operational technology] networks are somehow in danger. If you can say that the [indicators of compromise] we have look like a certain APT group, and that group does espionage and the country they belong to has really never shown an interest in sabotage, you can really calm them down. Espionage is bad, they can steal your intellectual property, but they probably won't destroy your OT networks… at least on purpose!

TR: How would you compare and contrast attribution done by the private sector and government agencies?

TS: I think they complement each other in methods and, most of all, in sources. I often say there are three main challenges in attribution: the first is data, the second is data, and the third is data. The sources are really the most important, and the difference is that law enforcement and government agencies can see some servers and request data from social media networks, stuff like that. On the other hand, the private sector has the treasure of telemetry—signals and feedback from their security products, of course. So they compliment each other, but any formal cooperation between the two is difficult because everyone has limitations on what data they can share. The private sector has its terms and conditions, and it can't share customer data with everyone else. The government has its own rules. But even if there isn't formal cooperation, the private and public sectors influence each other, and they're aware of what the other sector publishes.

> Only when [art historians] realize that the artist didn't paint all the paintings themselves —they had a workshop, where they may have made a sketch and had an assistant paint the background—then they were able to attribute it to a certain artist. That's the same problem we have with APT groups. They don't necessarily have a fixed set of people—some leave, some join, they might share malware or an idea for an exploit with another group—so we have to… understand that the whole landscape is more dynamic, like a workshop."

Another difference is that the private sector is really good at detecting new attacks and roughly clustering groups and malware. They often inform the agencies, which look further into certain things.

Analysts and security companies do something called continuous attribution… their main job is not really attribution, but protecting customers, writing signatures, improving their security products. But every now and then they stumble across something that can be useful for attribution. The law enforcement approach is different—they really have the explicit goal to do attribution. That's what they're paid for, that's what they're expected to do every day. They typically look at specific incidents, where you have a specific case, a certain victim, and they can't stop if they have a specific country—indictments can only be issued against individuals. So they have different goals and processes from the private sector.

TR: Do you think anything needs to change with how the public and private sector collaborate?

TS: Well in the best of all worlds, I think there are many opportunities to collaborate. But the main limitations are sharing data. So in the end, it works OK, there aren't hard boundaries between the public and private sectors… maybe you don't always exchange data on attribution, but you share things related to malware, infrastructure, and stuff like that. In the latest indictments, U.S. agencies have seemed to find a way to work together with some really good analysts from companies, but I don't know exactly what these arrangements look like, and I think it will be case-by-case for future collaboration.

TR: When it comes to attribution and naming, you have FireEye giving groups a number, CrowdStrike does animals, Microsoft does elements… Do you think we will ever get to the point where we'll have a unified system where we all know what each other is talking about?

TS: In my view, the clustering and definition of APT groups is where the magic happens and where the debate is going on. If you look at attribution, there's always a debate about which incidents belong to one group or another. So it's not just naming, it's really the underlying definitions.

I think my main point is that the whole concept of APT groups is just a simplification. It works as a model, just as the nuclear model works in physics, but everyone knows the reality is more complicated. When we talk about groups, it's not like a fixed set of people working together in an office for years and years. The reality is much more difficult. And for some of these setups, even if you had complete knowledge, I don't think it would be conceptually clear what the best or correct definition of groups would be. You have people that develop malware and then provide it to operators, and for the next job might take a piece of that and give it to someone else. The question then is, is this the same group? Groups also change over time—people age, they have families, they find different jobs or get promoted. It's really difficult to press this very complicated reality into APT groups, so we'll have to live with this challenge where every security company thinks their understanding of this group is the correct one and everyone else is wrong. This will certainly last for a long time, unless we figure out a better model or theoretical concept for how to describe these APT landscapes.

TR: How accurate do you think most attribution is?

TS: I think in general the industry is doing quite fine nowadays. They do have a certain set of methods and most people have a general understanding of how strong an attribution is, given the evidence. If you look at the indictments that the U.S. agencies released, in my view I trust them and I'm not aware of any big surprises. Everything the indictments said, it aligned with what the industry published beforehand.

Most companies agree at least on the attribution to countries, but it's different if you look at attributions to organizations or individuals. The indictments always contain something new because for the private sector it's really hard to attribute to individuals or organizations, and maybe it's not even their motivation to do this. So I think the state-of-the-art is that we're good on country-level attribution, but it's really hard on the organization or even individual-level.

TR: Do you think APT groups are getting better at covering their tracks?

TS: There have been several developments that complicate attribution. Groups outsource their tools and techniques, so it can be really hard to figure out if an attack is APT10 or a new group. It's certainly a problem. And the setup of groups is becoming more complicated, or maybe we only now understand the difficulty of grasping all these capabilities into one group.

On the one hand, as the state of attribution progresses, we have so much knowledge about groups and countries and what they're after—their goals and how they run their APT groups. This helps, since you don't have to start from scratch every time you want to make an attribution. But on the other hand, sometimes this preexisting knowledge is a problem, because we've been doing these attributions for ten years or longer, and the group that attacked ten years ago is probably not composed of the same people, setup, contractors, or maybe not even the same agency that runs it today. We need to find some rules or methods when it's better to just discard APT definitions and not try to match incidents to groups that you saw ten years ago.

TR: You've written a lot about false flags. What are the most common ones, and how do you identify them?

TS: False flags are a really fascinating concept, and not just because actors apply them so well. The existence of the concept itself is the problem. Anything that you find could, in general, be a false flag. When you look at the real false flags we've observed, most of them are quite similar. One thing that today most people wouldn't even call a false flag is basically domain registration data—it's always fake. Nobody really sees it as a false flag, but in general it is. Another common false flag is something written in a foreign language, but it's

often so badly done, because the people who plant these false flags don't speak the language. You find grammatical errors and bad Google translations, so it's really funny at times.

> Sometimes this preexisting knowledge is a problem, because we've been doing these attributions for ten years or longer, and the group that attacked ten years ago is probably not composed of the same people, setup, contractors, or maybe not even the same agency that runs it today. We need to find some rules or methods when it's better to just discard APT definitions and not try to match incidents to groups that you saw ten years ago."

But there can also be very complicated false flags. There was a famous and smart case of the Olympic Destroyer malware, where an analyst at Kaspersky found they copied a rich header to make it look like a known probably North Korean group. It was really complicated, but if you use some methods it's possible to infer that it may be a false flag. The malware isn't the only aspect you need to look into—you need to look at the infrastructure, domain registrations, telemetry, the stuff you might find on control servers. I'm not aware of threat groups that really consistently plant their false flags in all of these aspects of their campaigns —they might focus on one or two, often the malware itself because the industry often focuses on that.

TR: The state of attribution is something of an art right now. Can you talk about turning it into a science?

TS: My goal is really to establish attribution as an accepted discipline, like criminology. There's no guarantee that you can find a murderer or perpetrator, but there's a certain set of methods and processes and standards for how to present your results and establish confidence in hypotheses. I think it would be really beneficial if we could reach this level for attribution. Of course, this is a bit tempered by the fact that we're talking about espionage mostly, and so not all of the data can be transparent and public. It's challenging to make attribution a discipline, but I really think we should aim to do it. One of the steps is to codify methods and processes into a document, like a book. If you have these things documented, they don't have to start from scratch, they can start from there.

TR: You've mentioned how you see attribution as being similar to art history, too. Can you talk more about that?

TS: I think it's really odd that when you work in our field you always have to stay in threat intelligence mode. I really enjoy going into art museums, and when I'm there I can't help myself from thinking like a threat intelligence analyst. At one exhibition, I realized that art history has the same problem as attribution. You have some artifacts and you need to find who is behind them. They try to find the artist or painter, and we try to find the developer.

There are documented cases in art history where they try to figure out whether a certain painting was done by a certain artist or not, and they might say that a portion of a painting is not done as well as the artist would normally have done. So they're stuck. Only when they realize that the artist didn't paint all the paintings themselves—they had a workshop, where they may have made a sketch and had an assistant paint the background—then they were able to attribute it to a certain artist. That's the same problem we have with APT groups. They don't necessarily have a fixed set of people—some leave, some join, they might share malware or an idea for an exploit with another group—so we have to abandon our idea that everything is a fixed set of APT groups and understand that the whole landscape is more dynamic, like a workshop. It's a more difficult and complicated reality than just a single artist.

TR: What's a group that has been easy to track, in your opinion, and how does it compare to a group that has given you headaches?

TS: Everyone jokes about Gamaredon. Everyone can track them very easily, but nevertheless they are very active and they accomplish their goals. So in the end, it's not important whether you can be detected and attributed, but if you accomplish your job. They're very different from groups like APT29—they retool, get under the radar, and it can get hard to get back to tracking them. Once you do find them, you can unravel it and then you're fine again.

TR: How would you recommend a young analyst learn the tradecraft around attribution?

TS: I'm a fan of books. I had an AI background, and when I had my job interview with my first employer, I didn't know a lot of cybersecurity concepts. So I bought two books and went over them. So I think the first is getting good recommendations for general overview books, and then sticking with what you're most interested in and practicing it. Try to get feedback on your analysis. You'll fail, you'll get a lot of people telling you that you could do better, but the feedback is so valuable and there are many good people out there and most of them are willing to help and train and teach.

TR: How would you recommend sharing methods of tracking groups so as not to tip off adversaries now that in-person conferences aren't an option?

TS: I think this is quite established with the community, that we have some close groups and trusted corporations, and you should try to share your insights with these groups first who can then help the public and the wider community. Don't blast everything on Twitter. And also don't talk about attribution if you don't have first-hand knowledge and data on it.

TR: How did you get into this type of work?

TS: My background is in artificial intelligence and I also took a lot of psychology classes because I was really interested in how the human mind works, how human problem solving works, and whether you can simulate this with computers. Back when I was a student,

artificial intelligence was more about understanding the human brain and building something like Data—the android in Star Trek. Today it looks a lot more like data crunching…

My PhD was on learning opponent behavior in robotic soccer. I developed some algorithms and approaches to predict the moves of your opponent. Then I was hired to work on early warning systems—to predict the behavior of computer worms. Back then, 10-12 years ago, computer worms were a really big issue. Yes, we had WannaCry a few years ago, but before that computer worms were really dead. I think that's a lesson about how the threat landscape evolves in cycles, and I'm curious what we will talk about in five years.

TR: What motivated you to change your focus to attribution?

TS: That's a good question. I think it was really a byproduct of when I was an incident handler and having people constantly asking me who was behind an attack. It's also in line with my interests in the past. When I said I was interested in the human mind, I'm interested in how the opponent thinks and works, so I'm just trying to get as near to the people behind the attacks as possible.

Adam is the founding editor-in-chief of The Record by Recorded Future. He previously was the cybersecurity and privacy reporter for Protocol, and prior to that covered cybersecurity, AI, and other emerging technology for The Wall Street Journal.