Attribution: A Puzzle

By Paul Rascagneres

Published: 2020-08-13 · Archived: 2026-04-05 18:24:07 UTC

Thursday, August 13, 2020 09:08

By Martin Lee, Paul Rascagneres and Vitor Ventura.

Introduction

The attribution of cyber attacks is hard. It requires collecting diverse intelligence, analyzing it and deciding who is responsible. Rarely does the evidence available to researchers reach a level of proof that would be acceptable in a court of law.

Nevertheless, the private sector rises to the challenge to attempt to associate cyber attacks to threat actors using the intelligence available to them. This intelligence takes the form of open-source intelligence (OSINT), or analysis of the technical intelligence (TECHINT), possibly derived from proprietary data. Indicators in these sources tend to point toward a threat actor if they have used the same methods in the past, or reused infrastructure from previous attacks.

Intelligence agencies have additional sources of intelligence available to them that are not available to the private sector. The public saw a glimpse into this with a report that the Dutch agency AIVD compromised a security camera in the building used by APT29, an infamous threat actor. This allowed the Dutch Intelligence Agencies to provide vital intelligence regarding the activities of APT29 to their allies. Such intelligence is beyond the reach of private-sector researchers.

Intelligence agencies tend to be reserved, and publish relatively few articles that include attribution, at least in comparison to the private sector. Hence, when an intelligence agency, like the UK's National Cyber Security Centre (NCSC) directly attributed the WellMess malware to APT29 in a report endorsed by Canada's Communications Security Establishment (CSE), the U.S.'s National Security Agency (NSA) and Department of Homeland Security Cybersecurity and Infrastructure Security Agency (DHS CISA), you can expect that these agencies have solid evidence to back their claims.

Given this, it is interesting to examine the evidence available to us as a threat intelligence and security research group to support these conclusions. Attribution is typically not our goal. We aim to protect customers against threats, raise awareness of current threats, and support the security community. We recognize that we don't have the depth of visibility of an intelligence or law enforcement agency, but we do have access to a wealth of information, including open-source intelligence that helps us achieve our goals.

Infrastructure and TTPs sharing

The WellMess malware is an excellent example of how examination of infrastructure and the techniques used in an attack can lead to different conclusions. The Japanese national CERT named this malware in their July 2018 report. Two years later, the malware was used in attacks targeting COVID-19 vaccine research.

Written in Go, the malware has 32-bit and 64-bit variants, for both Linux (ELF) and Windows (PE) environments. The malware is known to use multiple protocols to conduct C2 communications including DNS, HTTP and HTTPS. The malware exfiltrates information regarding the system that it infected, before waiting for further instructions, which potentially include the execution of commands on the infected system and the exfiltration of additional data.

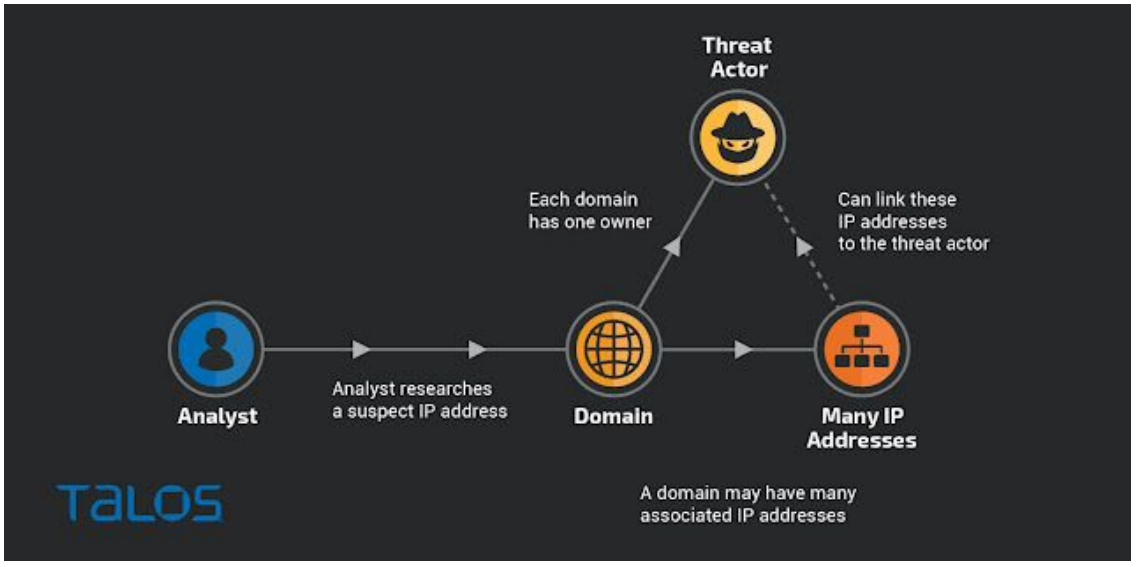
Attribution pivots

Infrastructure

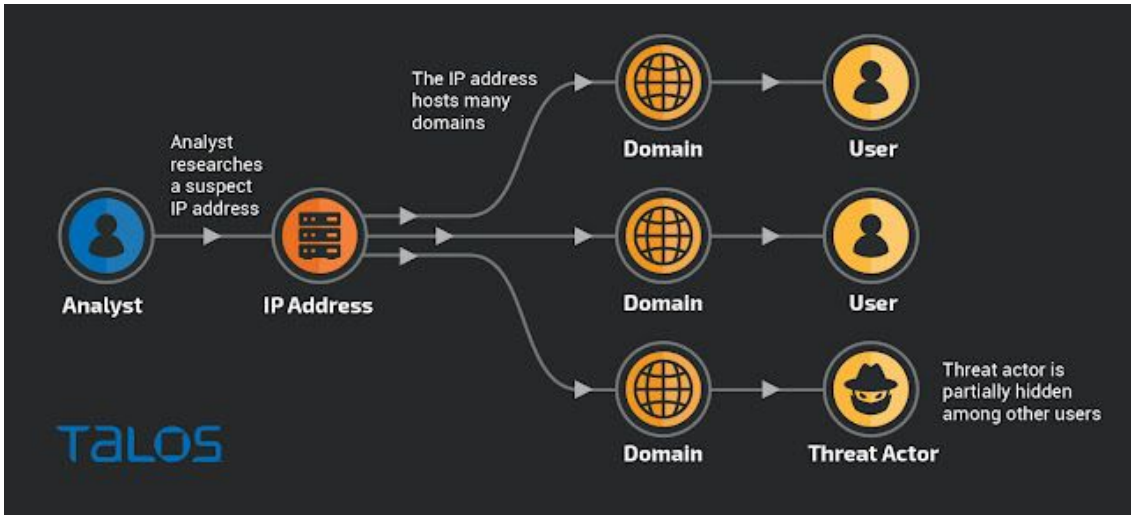
A common technique for linking separate campaigns is by pivoting on common infrastructure:

- Malware A uses infrastructure X.
- Malware B, associated with threat actor M, also uses infrastructure X.

Therefore, it is likely that malware A is also associated with threat actor M, linking the attacks through the shared infrastructure X



This technique can be used to investigate malicious IP addresses and domains. However, it's worth remembering the differences in ownership between domains and IP addresses. Domains are registered by their owners. This means that any IP address referenced by a domain is linked in some way to the domain owner.



On the other hand, IP addresses typically host services on behalf of many different organizations and individuals. The users of an IP address may be completely unknown to each other and have no awareness of the others' activities. Hence, while we can associate IP addresses with many domains, we can't be sure that the registered owners of these domains are linked without additional indicators.

We can investigate two samples of WellMess to check what we know about attribution.

Sample 0b8e6a11adaa3df120ec15846bb966d674724b6b92eae34d63b665e0698e0193 was submitted to VirusTotal on **May 23, 2018**. The C2 server of this sample is 45.123.190[.]168.

This IP address had two domains pointing to it over time:

- 2016-12-24 to 2019-12-04 layers[.]wincodec[.]com

- **2017-11-25 to 2018-11-18 onedrive-jp[.]com**
  We can link WellMess to **onedrive-jp[.]com**, because at the time the malware was submitted to VirusTotal, this domain was pointing to the IP address the malware was using for C2.

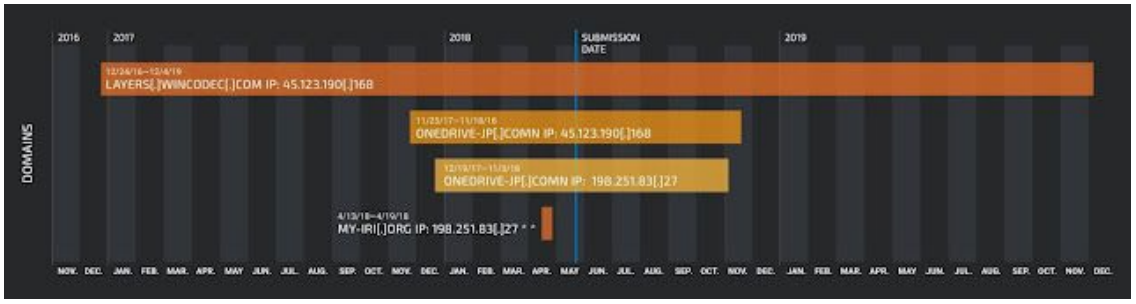
Let's check the history of this domain:

- 2020-07-17 to 2020-07-17 52.45.178[.]122
- 2018-11-22 to 2018-12-29 209.99.40[.]222
- 2018-11-21 to 2018-12-25 209.99.40[.]223
- 2017-11-25 to 2018-11-18 45.123.190[.]168
- **2017-12-19 to 2018-11-03 198.251.83[.]27**

Additional research showed that during an overlapping time frame, the IP address 198.251.83[.]27 is also linked to onedrive-jp[.]com.

Examining the history of this IP address, we find that the domain my-iri[.]org pointed to this IP address from **April 13 - 19, 2018**. During this time, the domain my-iri[.]org was mentioned by Microsoft as being associated with attacks against U.S. political institutions and attributed to APT28 — more details here and here.

These events in a timeline look like this:



The evidence is weak, but we have an association with APT28 rather than APT29 which is a different threat actor. APT28 and APT29 are usually considered to be different organisations within the same country. So here could be an alternative investigative path.

Unrelated note on this sample:

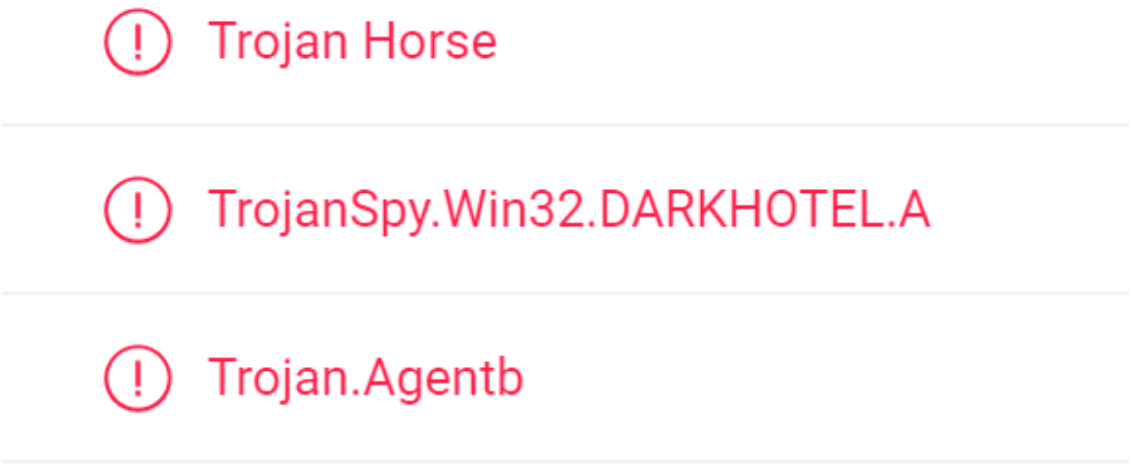
Another interesting element about this specific sample is the filename: QnapSSL. Qnap devices have x64 CPU and run on a Linux system. Potentially, this is an indicator that the malware may run on these devices.

TTPs

We can also examine the tactics, techniques & procedures (TTPs) of the threat actor, by identifying similarities in how campaigns are conducted.

Let's consider another WellMess sample: 65495d173e305625696051944a36a031ea94bb3a4f13034d8be740982bc4ab75.

The original name of the sample was "SangforUD.exe," the filename of the Sangfor VPN client. Examining VirusTotal shows that some of the detection engines detect this sample as "DarkHotel:"



The filename and detection name are reminiscent of the research report from the Chinese security firm Qihoo 360 on the DarkHotel attacks. Although the original report is no longer available, media coverage is still online here, and the press release from the date of discovery of the attack in April 2020 is available here.

In this attack, an attacker exploited a vulnerability in the Sangfor SSL VPN Server (which Qihoo alleges is DarkHotel) to replace the SangforUD.exe client binary with a trojanized version. When a client connects to the malicious server, it automatically downloaded and executed the new trojanized version of the VPN client.

This sample was submitted to VirusTotal on Jan. 21, 2020, a few months before the attack was discovered and published.

The DarkHotel APT group are a distinct threat actor known for targeting east Asia. Although this is a weak link, it creates another investigation path that, if properly corroborated, could lead to different attribution.

Analysing the evidence

The NCSC report clearly attributes the attack to APT29. We can't confirm or refute this conclusion, mainly because their intelligence is not publicly available and can be assumed to combine several different types of intelligence sources. Our own TECHINT-based research of the infrastructure indicates that WellMess might be associated with APT28. However, our TTP pivots suggest the malware could be linked to DarkHotel.

APT28 and APT29 are different organisations assumed to be running out of the same country, but DarkHotel is linked to a completely different country. It's possible that APT28 and APT29 may share the same geopolitical ambitions but given the publicly available information about these threat actors, it is very unlikely that they would share infrastructure. Furthermore, it is extremely unlikely that any of these two groups would collaborate with DarkHotel.

We can also use these observations to formulate some different hypotheses:

- The attribution concerning the Sangfor VPN servers hack may be incorrect. Was this an attack carried out by APT28 or APT29 rather than DarkHotel?
- Two different threat actors targeted the same VPN software at the same time by coincidence.
- Or, possibly, there is an unknown common factor between the threat actors that led to them targeting the same software.

To further complicate attribution, CoreSec360 published a post in Chinese about WellMess and the exploitation of VPN vulnerabilities here. In this post, the author attributes WellMess to an unknown actor they name APT-C-42.

Individually, each link appears feasible, but put into a wider picture they are incompatible with each other. Some of these links may be false flags, some may be due to coincidental factors. In any case, the best we can say is that attribution of this case is complex.

False flags

In some cases, false evidence is planted deliberately to confuse researchers. In acknowledging the existence of false flags, we must also admit it's possible researchers have misattributed attacks after being fooled by the threat actor.

One of the most egregious examples of false flags was that of Olympic Destroyer, the malware that disrupted the opening of the 2018 Winter Olympics. In this attack, the threat actor left clues in the malware that potentially implicated three different state-sponsored actors in carrying out the attack.

The logic of the wiper functionality of the malware was similar to that used by the Bluenoroff group, as was the file naming convention in a file system check. However, the technique of autonomous spread was like the NotPetya malware, attributed to APT28. Additional features in the code were also identified as having been used by APT3 and APT10 in the past.

It is likely that the threat actor understood how analysts arrive at conclusions of attribution and wanted to include artefacts to lead analysts astray. We have already discussed these in detail in this blog post.

In this case, attribution via TECHINT was not possible. The pivot points that might be used by researchers had been deliberately falsified. However, additional intelligence was almost certainly available to intelligence agencies. In unofficial briefings, US intelligence personnel reportedly linked the Olympic Destroyer attack to the Sandworm group. Subsequently, EU sanctions listed Olympic Destroyer as an alias of the Sandworm group, supporting the attribution.

Cases such as these make a good point for analysts not to rely on TTPs or code analysis alone. Additional intelligence is required — usually of the level that only intelligence agencies have — to untangle the threads. Understandably, intelligence agencies are often reticent to disclose too much information or provide proof that can be externally verified.

Analysts must be open to the idea that any clues they find might have been planted deliberately, and to keep an open mind until the full picture can be assembled.

Software engineers like to reuse tried and trusted code in their projects, including code that they are familiar with, and that they know works increases productivity and helps ensure that their project will be delivered on time and within budget. Threat actors are no different.

We follow the development programs of threat actors and observe any features they add to their malware. The Bisonal malware associated with the Tonto Team which we published recently is a good example of this. We can follow the software development program of the threat actor over 10 years.

However, there are pitfalls in this approach. In June 2020, Unit42 identified the ACIDBox malware. Unit42 was clear in their conclusion that they cannot attribute ACIDBox to the threat actor Turla.

A few days later, a prolific and respected security researcher published a blog highlighting code similarities between ACIDBox and Turla Nautilus samples here. This was based on a piece of code shared between the two malware that although small, seemed unique enough to be used for attribution. Without claiming absolute certainty, the researcher posted their findings for the community to vet. One eagle-eyed researcher on Twitter, @TheEnergyStory, pointed out the shared code is, in fact, compiled Mbed TLS implementation code.

Like many software engineers, the threat actor chose to include a common library in their project to provide functionality that they didn't want to redevelop themselves.

Finding common code between malware doesn't necessarily show that the samples are related. Third-party libraries can be included independently in different samples without the samples being related. Of course, finding a shared library doesn't discount the possibility that two samples are related, but it doesn't contribute to the debate.

This example illustrates another important point. Sharing information with the community of security researchers and inviting feedback is vital. Within a few hours of the original publication, a third-party researcher identified the source of the shared code and explained its inclusion. Thus allowing the original researcher to update their research with the new information.

Conclusion

Attribution is complex and difficult at the best of times. Traditional intelligence agencies rely on a variety of intelligence sources to arrive at a conclusion. This type of information is often not available to private companies, who tend to make conclusions solely from TECHINT which can be problematic. Once we take into consideration that threat actors may be purposefully conducting false-flag operations to deliberately confuse analysts, attribution becomes even more difficult.

The challenge of attribution, in particular with the complication of false-flag activity, means that some of the industry's chains of associations upon which allegations of attribution are made may be incorrect. Private-sector researchers, without the support of traditional intelligence capability, should be cognizant of these risks.

When a set of some of the world's most well-known intelligence agencies attribute an attack to a threat actor, this attribution cannot be discounted off hand — even if they do not provide the community with all the information that supports their attribution. With the information and techniques available to us, we cannot confirm such an attribution. This is not to say that we refute the attribution, but that we are missing the intelligence or logic that was used to come to such a conclusion.

Attribution is as much of a science of collecting verifiable information as it is an art of assembling a hypothesis and being aware of the information which is missing to support that hypothesis. Hence, allegations of attribution should be approached with an open mind but being aware that the body making the allegation has more information than they are letting be made public.
