Skopik and Pahi *Cybersecurity* (2020) 3:8
https://doi.org/10.1186/s42400-020-00048-4

Cybersecurity

RESEARCH Open Access

Under false flag: using technical artifacts for cyber attack attribution

Florian Skopik* and Timea Pahi

Abstract

The attribution of cyber attacks is often neglected. The consensus still is that little can be done to prosecute the perpetrators – and unfortunately, this might be right in many cases. What is however only of limited interest for the private industry is in the center of interest for nation states. Investigating if an attack was carried out in the name of a nation state is a crucial task for secret services. Many methods, tools and processes exist for network- and computer forensics that allow the collection of traces and evidences. They are the basis to associate adversarial actions to threat actors. However, a serious problem which has not got the appropriate attention from research yet, are false flag campaigns, cyber attacks which apply covert tactics to deceive or misguide attribution attempts – either to hide traces or to blame others. In this paper we provide an overview of prominent attack techniques along the cyber kill chain. We investigate traces left by attack techniques and which questions in course of the attribution process are answered by investigating these traces. Eventually, we assess how easily traces can be spoofed and rate their relevancy with respect to identifying false flag campaigns.

Keywords: Actor attribution, Advanced persistent threats, Technical indicators, False flag campaigns

Introduction

A false flag in the cyber domain is significantly different and much easier to carry out than in the physical world (Goodman 2010). Cyber false flags refer to tactics applied by cunning perpetrators in covert cyber attacks to deceive or misguide attribution attempts including the attacker's origin, identity, movement, and exploitation. It is typically very hard to conclusively attribute cyber attacks to their perpetrators and misdirection tactics can cause misattribution (permitting response and counterattack, which can lead to retaliation against the wrong party (Wheeler and Larsen 2003; Philbin 2013; Harrington 2014)).

False flag operations have long existed in the physical world (Kearns et al. 2014), a tactic used to make an operation appear to have been planned and executed by someone other than the real perpetrator. Digging a little deeper into the concept of a false flag operation shows that the intent of the actor behind the operation is to do one of two things: (1) let a third party take the blame for malicious actions they did not carry out, or (2) hide own malicious actions behind someone else (Morgan and Kelly 2019).

Cyber attribution (Rid and Buchanan 2015) is not an easy task, and the existence of false flags make the situation even worse. However, attribution is a crucial task of nation states to carefully distinguish between the acts of criminal organizations and the acts of war of nation states. Wrong attribution can have devastating consequences, which leaves zero tolerance for failures.

It is therefore of utmost importance to get it right. An important prerequisite is to know about common attack tools and techniques and understand which traces (typically artifacts) they potentially leave in a victim's infrastructure (or even elsewhere, like at the premises of cloud providers and other external parties). Starting from there, it is further important to understand what questions during the investigations of an incident can be answered by studying these artifacts. And eventually, it is key to understand how reliable the conclusions based on the

*Correspondence: florian.skopik@ait.ac.at
Center for Digital Safety and Security, AIT Austrian Institute of Technology, Austria, Giefinggasse 4, Vienna, Austria

investigation of certain artifacts are. In fact, some artifacts could be spoofed, certain traces faked to point at other parties than the real adversary. In a complex technical system this is realistic; however, the complexity of today's systems does it not only make hard for investigators to fetch the relevant data to achieve vital insights, but it also makes it hard for cunning attackers to consistently carry out false flags campaigns. This benefits the investigators. Eventually we must acknowledge that cyber attack attribution is challenging, but nevertheless important – and when done, it allows no room for mistakes. In this paper we take a closer look into this dilemma.

The contributions of this paper are:

- **Short survey on attack techniques:** We outline common cyber weapons and attack techniques along the cyber kill chain. Every applied technique leaves different kinds of traces. In the early phase of an attack this might be information related to domain registration or bitcoin transactions used to rent a service in the dark net. During the actual attack, entries in log files of exploited machines may point to applied exploits or C&C traffic.
- **Relevant artifacts:** For each phase in the kill chain different questions may be asked in the forensic investigations. We highlight some relevant example questions and connect them to typical artifacts produced through the application of aforementioned attack techniques. Profound knowledge about artifacts that carry information about an attack aids the attribution process.
- **Attribution process and issues with false flags:** We outline what information is potentially derived during the attribution process through investigating the collected artifacts and discuss the issues with identifying false flags. Especially, we take a closer look into how easily the discovered artifacts can be spoofed by attackers to disguise traces and discuss their trustworthiness in course of an illustrative attribution scenario.

The remainder of this paper is organized as follows. "Related work" section provides an overview of related work, mainly around attributing APT attacks and discussions on false flag campaigns. "Common cyber weapons in the kill chain" section surveys the most prominent attack techniques along the cyber kill chain. "Artifacts for the attribution process" section discusses the critical questions being asked in course of investigations for the different phases of an APT attack, and highlights information sources and artifacts that may help attributing an attack correctly. "Attribution and false flags" section elaborates on the actual attribution process and the potential of spoofed artifacts in course of false flag campaigns. It further discusses an illustrative attribution scenario and demonstrates the application of methods presented in this paper. Finally, "Conclusion and future work" section concludes the paper.

Related work

Advanced persistent threats

APTs (Tankard 2011) have been studied extensively since the appearance of Stuxnet (Falliere et al. 2011). Some recent APTs, which we also investigate further in this paper (see "Attribution and false flags" section), are the Narwhal Spider APT, Grey Energy, Pro-Syrian Government Hackers, Octopus APT, and Darkhotel APT. These APTs represent a broad view on the different modes and ways attacker groups may work. They cover also the latest TTPs, the Narwhal Spider group uses for example a combination of steganography and malicious PowerShell (Beatty 2019). Grey Energy is more stealthy and sophisticated, as his ancestor Black Energy resulting in the first publicly-reported blackout caused by cyber operation (Cherepanov 2018). The Pro-Syrian Government Hackers differs from the other groups since they are directly political active. They spread a fake security tool in order to monitor and detect anti-government content and political-not-correct persons (Galperin and Marquis-Boire 2012). The Octopus APT used one of the most popular Social Media application for spreading fake news to incite unrest. Darkhotel belongs to the full-scale surveillance campaigns. Darkhotel had been active for seven years in a number of luxury Asian hotels (Woodier and Zingerle 2019). These five different cases form the basis of the survey about trustworthy artifact types in this work.

In order to allow for a structured analysis, numerous models to distinguish the phases of APTs have been proposed, of which some of the popular ones are the Lockheed Martin cyber kill chain (Yadav and Rao 2015; Hutchins et al. 2011) and the ATT&CK framework from MITRE (MITRE 2019). We use these models to establish an approach for a structured analysis of artifacts created in course of cyber attacks that possibly aid the attribution process. In particular, we use the categories of known attack vectors to survey relevant artifacts usable for attribution purposes.

Cyber attack attribution

A prerequisite of cyber attribution is to discover the applied techniques, tools and procedures (TTPs). Based on that, the further goal is to identify the source of certain attacks that leads to the threat actor. Both topics, cyber attack investigation (i.e., get to know what happened) and threat actor attribution (i.e., get to know who did it) aims to serve as a basis for actions in law enforcement and national security (such as cyber war or terrorism). There is a wide range of literature on this topic with different

approaches. It is often a mix of technical attack analysis and threat actor profiling, that sometimes leads to confusion. Numerous works have investigated the "art" of cyber attribution (Tran 2018; Tsagourias and Farrell 2018; Rid and Buchanan 2015; Tsagourias 2012; Brenner 2006). For instance (Tsagourias and Farrell 2018) outlines what data and information is most helpful in the attribution process to identify the perpetrator. They highlight that not only malware samples and their specific properties (such as compiler settings, language settings, certain re-occurring patterns and the like) are useful, but also information available outside the actually attacked infrastructure, including data on the command & control infrastructure. The latter includes knowledge of what IP addresses have been used, what domain names and registration information of these domains. This boils down to payment information for the referenced domains. Eventually (Tsagourias and Farrell 2018) comes up with a categorization of data which distinguishes between physical persons, virtual personas, campaigns and infrastructure and tools. Paper (Li 2014) extends this view specifically by DNS patterns, especially the association of domain names with IP addresses, which even extends the passive DNS (Bilge et al. 2011) project.

Attribution models

One of the best known models is the Q-Model by Rid & Buchanan (Rid and Buchanan 2015). They are looking for an answer to the question, whether attribution is a technical problem or not. Their model introduces multiple levels: strategic, operational, tactical and technical and communication, and several roles: from forensic investigators through national security officers to political leaders. The Q-Model helps analysts to ask the full range of relevant questions and put the investigation into context. It integrates both technical and non-technical information into competing hypotheses. This includes asking more challenging questions on different levels. The study 'Role and Challenges for Sufficient Cyber-Attack Attribution' (Hunker et al. 2008) summarizes the political, legal and technical challenges. A detailed description of legal issues is available in 'The Law of Attribution: Rules for Attributing the Source of a Cyber-Attack' (Tran 2018).

The Q-model (Rid and Buchanan 2015) is of particular interest to us as it already itemizes the relevant questions to be answered in course of investigations. However, while this model provides a structured view on the questions it does not elaborate on how to answer them. We pick up from there and try to provide – at least for some of the questions – answers on which artifacts are the most relevant ones and which traces possibly lead to the attackers. In terms of attribution it is also important to note that the goal is usually not to associate individuals to cyber attacks but whole groups (Lemay et al. 2018). Often the team members of these groups change and can hardly be identified, but it is important to understand the motives of a group and whether governments can be held accountable for their actions (Tsagourias 2012).

Regarding threat actor attribution, the Hacker Profiling Projects provides one of the most complex models (Chiesa et al. 2008). The research has four principal points of view: technological, social, psychological and criminological. Their profiling methodology contains the 4Ws: who, where, when, why. The resulting hacker profiles contain the following categories: Wanna Be, Script Kiddie, Cracker, Ethical Hacker, Skilled Hacker, Cyber-Warrior, Industrial Spy, Government agent, and Military hacker. The applied correlation standards cover the following aspects for each profile: modus operandi, lone hacker or group, selected targets, hacking career, hacker's ethics, crashed or damaged systems, perception of illegality and effect of laws. The study of PWC developed other profile categories: governments, criminals, hacktivist. They distinguish the perpetrators on the motivation and the technical origin of the attack: cyber crime, cyber activism or cyber warfare. The study 'Cyber Attribution Using Unclassified Dat' (Public-Private Analytic Exchange Program Team 2016) focuses on the Diamond Model and on accountability for investigation and prosecution based on cyber attribution. The results show that the cooperation between distinct communities (law enforcement, intelligence community, industry) is required for attribution, and there are no standardized tools in use. The Diamond Model appears again in the research of intrusion analysis (Caltagirone et al. 2013).

The Cyber Attribution Model (CAM) (Pahi and Skopik 2019), proposed as a response to the lack of other models to support the full attribution process, initially separates the two corner stones cyber attack analysis and threat actor profiling and brings them together in the attribution phase. In this paper we take a closer look into how this model can be populated with reliable technical data and outline the attribution process in course of an illustrative example.

False flags

The challenges related to false flags have been investigated before (Pihelgas 2015). Especially (Bartholomew and Guerrero-Saade 2016) takes a closer look into the problem and also surveys noteworthy actors in the field. Cyber Information sharing (Skopik et al. 2016; Wagner et al. 2016) is a common means to gather important data to aid the attribution process, for instance, information on attackers, their capabilities, used TTPs and so on.

It is important to note that attack detection is very much different from attribution, in that sense that not all sources (Zimmermann 2014; MITRE 2019) relevant to detect attacks (Caltagirone et al. 2013) are also appropriate for

attribution. In contrast to pure detection of attacks, attribution sets its focus to relating actions to actors. A particularly important piece are therefore actor profiles (Chiesa et al. 2008) used to correlate identified actions to known capabilities and applied tactics and techniques of actor groups. The main problem of the attribution is a potential misattribution. Since the perpetrator attempt to cover up their tracks through a mixture of evasiveness, deception, and destruction of records or though false flags. This issue is well known among agencies and secret services. The NSA and NCSC released for instance a joint advisory APT group to avoid possible misattribution [1]. The information sharing, especially threat intelligence, is one essential aspect to detect false falgs. There are some initiatives worldwide, such as the Cybersecurity Information Sharing Act in the USA [2], Cyber Security Information Sharing of ENISA [3], or Cyber Information Exchange in the NATO [4]. The NATO Cooperative Cyber Defence Centre of Excellence is also aware of the issue and developed own definitions for false-flag, no-flag cyber operations (Pihelgas 2015) and influence cyber operations (Brangetto and Veenendaal 2016). Influence cyber operations are designed to influence the behaviour of a target audience. These false flag operations are part of the hybrid threats today, such as Russian aggression against Ukraine in 2014 and its intervention in Syria in 2015. One of the latest false flag operation is the TV5Hack (see "Illustrative application of CAM to identify false flags" section).

Common cyber weapons in the kill chain

Numerous initiatives attempt to structure the different attack techniques used in complex multi-stage APT attacks, including the Lockheed Martin cyber kill chain (Yadav and Rao 2015) and the ATT&CK framework from MITRE (MITRE 2019). While the latter is quite tool-centric and details the different stages of carrying out the actual attack, the first one also considers the preparation phases of an attack, including reconnaissance. We argue that even in these phases, when attackers prepare for a complex attack, traces may be left, such as the attempts to buy zero day exploits in the dark net, excessive scanning activities, social engineering attempts and the like. We therefore survey most common techniques along the cyber kill chain, which consists of the following phases:

1 **Reconnaissance:** First, attackers probe for a weakness. This might include harvesting login credentials or information useful in a phishing attack.
2 **Weaponization:** This phase deals with building a deliverable payload, usually using an exploit within the target system and a backdoor to maintain long-term access.
3 **Delivery:** Then the attackers send the weaponized bundle to the victim, e.g., a malicious link in a legitimate-looking e-mail.
4 **Exploitation:** Upon successful delivery, the initial intrusion takes place by executing code on the victim's system. Installing a backdoor allows attackers to maintain permanent access, even if the initially exploited vulnerability is fixed.
5 **Lateral Movement and Installation:** Then, the attackers pivot through the system to find valuable resources. Once a particular target machine has been spotted, more sophisticated malware is dropped.
6 **Command & Control:** This malware allows to create a channel where the attacker can control a system remotely.
7 **Actions:** Eventually the attackers remotely carry out their goal, e.g., to take over control or exfiltrate data.

In this section we survey frequently applied techniques to run through the cyber kill chain. In the next section, we are going to investigate which actual traces the application of these techniques leaves and what artifacts need to be collected and studied to aid the attribution process. Notice that this list is far from being complete as this would certainly go beyond the scope of this paper. However, we surveyed prominent APT cases (cf. "Related work" section) from the past years and highlight the most frequently used techniques.

Step 1: reconnaissance

The execution of the reconnaissance phase is individually shaped to the attack and unique in every case. Here, the attacker attempts to acquire the required knowledge to better understand the victim's infrastructure, discover weaknesses and vulnerabilities, including non-technical ones, and gather further knowledge that helps to find a path into the target organization. The latter could be quite tricky and often does not only include the actual target, but also vertical organizations, such as suppliers, customers and partner organizations. Common attack techniques are:

- **Open Source Intelligence (OSINT) analysis:** analysis of publicly available profiles at social network sites, private homepages, company pages, job advertisements, public relations material, Q&A forums etc. (Hulnick 2010)
- **Social engineering:** spear phishing (Hadnagy 2010) to gather further information, baiting, i.e., the distribution of infected USB sticks (not for intrusion

[1] https://www.us-cert.gov/ncas/current-activity/2019/10/21/nsa-and-ncsc-release-joint-advisory-turla-group-activity
[2] https://www.cisa.gov/about-cisa
[3] https://www.enisa.europa.eu/publications/cybersecurity-information-sharing
[4] https://ccdcoe.org/library/publications/whole-of-government-cyber-information-sharing

but information gathering), tailgating, e.g., testing for physical access as fake client (Krombholz et al. 2015).
- **Passive scanning:** intercepting traffic from/to the target network (Bartlett et al. 2007), elaborating WiFi ranges and settings.
- **Active network scanning and enumeration:** Banner grabbing (Kang et al. 2017) of publicly facing services for subsequent vulnerability analysis, probing of (web) server capabilities, port scans to discover firewall settings for certain IP ranges (Lyon 2009).

Step 2: weaponization

The weaponization mainly consists of building a deliverable payload that applies an exploit to place a backdoor. It further may include steps to build up a working external infrastructure, i.e., acquiring domains later being used to create command and control (C&C) capabilities. Common attack techniques are:

- **Set up a staged lab environment** built to research vulnerabilities and evaluate numerous attack techniques (Peterson 2013).
- **Evaluate standard-payloads**, e.g. reverse shell (Foster et al. 2015), keylogger (Tuli and Sahu 2013), snapshot tools together with well-known demo code for vulnerabilities (depending on the target requirements and attacker capabilities).
- **Research zero days exploits** (Bilge and Dumitraş 2012), either developed individually (which is resource-intensive and requires access to copies/clones of the target equipment) or purchased on the black market (requires access to these sources; payment usually via electronic currencies). With respect to the latter, we must distinguish between buying bare proof-of-concepts which are further modified and compiled by the attacker v.s. full-fledged exploitation software.
- **Crafting a remote access trojan** (Haagman and Ghavalas 2005), typically bundled with an exploit to be dropped into the target environment.
- **Survey different delivery techniques**, e.g., infected pdf, macros in docx, xlsx, scr files etc.
- **Preparation of external infrastructures**, such as registration (or even reuse) of domain names, botnets (Li et al. 2009) and cloud services to establish C&C capabilities compatible to the constructed payload.

Step 3: delivery

After crafting the payload, the attackers either send the weaponized bundle to the victim, e.g., a malicious link in a legitimate-looking e-mail, or deploy it somewhere, where the victim is likely to pass by (e.g., a manipulated ad banner wich exploits a JavaScript vulnerability). Notice, sometimes the payload is not directly delivered to the target organization but rather to an associated organization, such as a partner company or sub contractor, to sneak into the supply chain. Common attack techniques are:

- **(Spear) Phishing** (Krombholz et al. 2015) tries to persuade employees to act in the interest of the attackers, e.g. click a link or download a software.
- **Infection of resources or services of an associated company** or supplier, including subcontractors, partner companies or customers and exploit the trust relation between those entities and the actual target organization.
- **Watering hole attacks** (Krombholz et al. 2015) try to push malware onto a victim's computer and exploit browser vulnerabilities, once they visit a particular Internet site.
- **WiFi and extended network infrastructure** are means to get access to an otherwise well secured network segment.
- **Physical access** to manipulate equipment directly or to drop malware via social engineering, e.g. baiting (Hadnagy 2010).
- **"Traditional" criminal activities**, such as bribery or blackmailing are old, but nevertheless an effective means to get initial access.

Step 4: exploitation

After the successful delivery of the payload, the initial intrusion takes place by executing code on the victim's system. Installing a backdoor allows attackers to maintain permanent access, even if the initially exploited vulnerability is fixed. Common attack techniques are:

- **Exploits**, such as the initial exploitation of a (Browser) vulnerability, a vulnerable plugin (for Java, Flash etc.), or a product used to view a manipulated file, e.g., MS Office, Adobe Acrobat and the like.
- **Backdoors**, including the installation of a remote access trojan/tool (RAT) (Haagman and Ghavalas 2005) which usually connects in reverse mode to the C&C server to pull commands from.
- **Cross-Site-Scripting (XSS)** (Vogt et al. 2007) exploited by specifically drafter requests to Web servers, may be used to bypass certain access controls and gain access to machines.

Step 5 to step 7: installation, command&Control, actions

Attack techniques employed in the later stages of the kill chain are quite diverse and manifold. MITRE did an excellent work summing them all up in their ATT&CK framework (Strom et al. 2017; MITRE 2019). Thus, we will not go into further details of the single techniques applied once an attacker got initial access, but rather sum up their underlying tactics (cf. overview in Fig. 1).

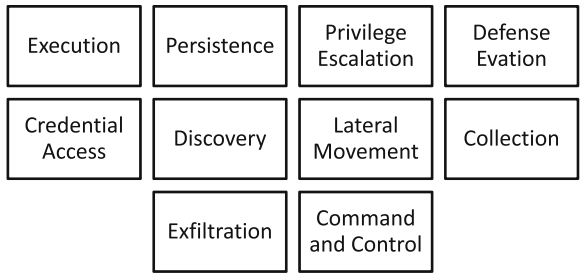


Fig. 1 Adversary enterprise tactics according to MITRE's ATT&CK framework

Execution: The execution tactic represents techniques that result in execution of adversary-controlled code on a local or remote system. This tactic is often used in conjunction with initial access as the means of executing code once access is obtained, and lateral movement to expand access to remote systems on a network.

Persistence: Persistence is any access, action, or configuration change to a system that gives an adversary a persistent presence on that system. Adversaries will often need to maintain access to systems through interruptions such as system restarts, loss of credentials, or other failures that would require a remote access tool to restart or alternate backdoor for them to regain access.

Privilege Escalation: Privilege escalation is the result of actions that allows an adversary to obtain a higher level of permissions on a system or network. Certain tools or actions require a higher level of privilege to work and are likely necessary at many points throughout an operation. Adversaries can enter a system with unprivileged access and must take advantage of a system weakness to obtain local administrator or SYSTEM/root level privileges. A user account with administrator-like access can also be used. User accounts with permissions to access specific systems or perform specific functions necessary for adversaries to achieve their objective may also be considered an escalation of privilege.

Defense Evasion: Defense evasion consists of techniques an adversary may use to evade detection or avoid other defenses. Sometimes these actions are the same as or variations of techniques in other categories that have the added benefit of subverting a particular defense or mitigation. Defense evasion may be considered a set of attributes the adversary applies to all other phases of the operation.

Credential Access: Credential access represents techniques resulting in access to or control over system, domain, or service credentials that are used within an enterprise environment. Adversaries will likely attempt to obtain legitimate credentials from users or administrator accounts (local system administrator or domain users with administrator access) to use within the network. This allows the adversary to assume the identity of the account, with all of that account's permissions on the system and network, and makes it harder for defenders to detect the adversary. With sufficient access within a network, an adversary can create accounts for later use within the environment.

Discovery: Discovery consists of techniques that allow the adversary to gain knowledge about the system and internal network. When adversaries gain access to a new system, they must orient themselves to what they now have control of and what benefits operating from that system give to their current objective or overall goals during the intrusion. The operating system provides many native tools that aid in this post-compromise information-gathering phase.

Lateral Movement: Lateral movement consists of techniques that enable an adversary to access and control remote systems on a network and could, but does not necessarily, include execution of tools on remote systems. The lateral movement techniques could allow an adversary to gather information from a system without needing additional tools, such as a remote access tool.

Collection: Collection consists of techniques used to identify and gather information, such as sensitive files, from a target network prior to exfiltration. This category also covers locations on a system or network where the adversary may look for information to exfiltrate.

Exfiltration: Exfiltration refers to techniques and attributes that result or aid in the adversary removing files and information from a target network. This category also covers locations on a system or network where the adversary may look for information to exfiltrate.

Command and Control: The command and control tactic represents how adversaries communicate with systems under their control within a target network. There are many ways an adversary can establish command and control with various levels of covertness, depending on system configuration and network topology.

Artifacts for the attribution process

Based on the overlook of attack techniques in "Common cyber weapons in the kill chain" section, we are going to investigate which types of artifacts are left by applying these techniques on the victim's computers, which potentially aid the attribution process. Knowing, which traces attackers leave help to better understand which of them are quite unique and which can be spoofed or tampered with easily.

Reconnaissance artifacts

When investigating an attack, typical questions concerning its very beginning are centered on whether the attackers already left traces in form of suspicious behavior. This

might include active scanning attempts on the network layer, profile mining activities on social networking sites, phishing for information gathering, faked visits including job interviews, password brute force attacks on externally facing services, e.g., Webmail and so on. Another important question to answer is whether attackers prepared for a direct attack or an indirect one, i.e., using a contractor or customer, to get into the target infrastructure. Typical artifacts and hints that may aid the attribution process:

- **Perimeter monitoring logs:** Logs can reveal scanning attempts (Kaushik et al. 2010) on layer 3 and 4 from a certain IP range and/or botnet which significantly exceed the usual noise (Li et al. 2009).
- **Social networking statistics:** Repeated visits of social network profiles (e.g., on LinkedIn) used to view a number of target profiles (correlation of viewer information across company staff).
- **Identities:** Numerous identities and personas used e.g., on social networking profiles to connect to company staff, phishing e-mail sender, fake identities to register domains, certificates used in communication (S/MIME, SSL), or to perform electronic payments (e.g. via bitcoins) (Liao et al. 2016), etc.
- **Spelling:** Typical or reoccurring spelling errors in phishing mails for information gathering; derived hints of the author's mother tongue due to specific language mistakes (Afroz et al. 2012).
- **Domains and DNS:** Names and registration information of domains referred to in phishing mails; especially in case of reused domains (Passive DNS (Bilge et al. 2011)).

Weaponization artifacts

Insights from collected traces of the weaponization phase (i.e., the preparation of the first step of the attack) must not be underestimated. Eventually, getting a foot into the door usually turns out as one of the most critical parts of a cyber operation. Main questions here are centered on the initial penetration technique, i.e., how complex it was and how much effort it was for the attackers. Especially of interest is whether known vulnerabilities were exploited and known tools used, or entirely novel ones, probably developed specifically for the investigated attack. The two ends of the same scale are on the one side the application of known/reused "hacking tools" for well-known technologies and on the other side the usage of newly written exploits for quite exotic software. This allows drawing interesting conclusions about the capabilities and resources of the attackers. Typical artifacts and hints that may aid the attribution process:

- **Malware analysis results:** Forensic analysis of the initially deployed malware and/or RAT (Alperovitch and et al. 2011) respectively, downloaded to or pushed towards a victim's computer (see Section on Exploitation artifacts for details).
- **Urls and IDs:** Download locations, URLs and IDs embedded in the malware.

Delivery artifacts

In this phase the constructed weapon for the initial infection is being delivered. The focus of the investigation is therefore on the delivery path of Malware (e-Mail, IM, Internet forum, injection, drive-by download etc.) and connected traces, such as ids, addresses, names and urls. The latter provide a convenient means to associate new attacks with old ones. Typical artifacts and hints that may aid the attribution process:

- **Phishing e-mails:** One of the most common delivery methods are still phishing e-mails (Krombholz et al. 2015; Hadnagy 2010). Information about the e-mail address (auto generated v.s. manually generated, legitimate v.s. spoofed address, language of address and domain), the e-mail content (grammar/language mistakes, salutation and sophistication level) provide hints on the modus operandi of the attacker. Of particular interest is if and how much insider knowledge was used, e.g., to craft a convincing mail that appears to come from a customer or contractor.
- **Identities, pseudonyms, and personas:** These might reveal hints on the cultural background of the attackers.
- **Delivery method and path:** Apart from e-mail, there are many other delivery methods, electronic ones (instant messenger, social networks, Web-based drive-by-download (Cova et al. 2010)) and physical ones (e.g., baiting (Bowen et al. 2009)), which might be used in different forms (e.g., embedded in another file)
- **Technical vulnerability exploited for delivery:** The delivered malware usually uses technical vulnerabilities (Kumar et al. 2006) to establish a foothold. However, also for the delivery, a technical vulnerability might be used, e.g., an injection attack.

Exploitation artifacts

Once malware was successfully delivered, it will be executed to establish foothold in the target infrastructure. Typically, some form of RAT is dropped and installed using a vulnerability. Questions are centered on this aspect, e.g., What vulnerability was exploited? How hard was it to exploit? How much background or even insider knowledge was likely required? How sophisticated

was the exploit? Where multiple techniques applied? Another aspect that tells a lot about the attacker is whether the malware hit a carefully selected machine or the infection rather took place by chance. Indicators for a targeted attack are when a rare vulnerability was exploited, or when the exploit was shaped to the environment, such as the rights of the logged-in user. Typical artifacts and hints that may aid the attribution process:

- **Malware attributes:** A sample of the malware allows detailed forensic investigation (Ligh et al. 2010). This might reveal common strings, such as variable names (depending on the programming language), embedded urls and ids, and the actual coding style. Deeper analysis might even reveal differences in coding style (e.g., how error handling is performed) and therefore allow assumptions on the circumstances this malware was produced in (e.g., division of labor, etc.). Depending on the programming language and the compiler further information such as personalized strings (e.g., paths on the development machine including user names) might be present. This is especially the case with interpreter languages, such as Python.
- **Malware metadata:** This includes all information on the circumstances under which Malware was constructed and which can be extracted in many cases from a sample. For instance, compilation time and compiler version used (which might point to applied tool chains), the programming language used, language settings and region settings of the development platform.
- **Malware design:** Deeper investigations may reveal information on the malware design, e.g. whether it is a novel monolithic piece of code or rather a compilation of different modules (e.g., initial exploit, reverse shell (Foster et al. 2015), keylogger (Tuli and Sahu 2013), cryptolocker (Liao et al. 2016) etc. – perhaps purchased from numerous different sources). Furthermore, the application of anti-forensic and obfuscation techniques (Harris 2006) is a good indicator for the sophistication level.
- **Malware functionality:** The functionality corresponds to the design, but the actual question here is, whether the initial malware was a simple RAT or something more sophisticated, e.g., was able to use alternative infection vectors, disguise its operations and so on. Furthermore, features and settings specifically shaped to the victim' machine(s) and their environment allows conclusions on how target-oriented the attack was (Jarvis and Macdonald 2015).

Installation artifacts

After the initial infection, attackers and/or their malware respectively will start to scan the network and locate the actual resources of interest (lateral movement). This phase usually involves quite complex actions, which are carried out individually depending on the attacker and on the victim. Therefore, almost all relevant questions in an investigation concerning this phase focus on the applied tactics, techniques and procedures (TTPs), which are a vital means to associate observed behavior to concrete attackers. Notice that from behavior a lot can be inferred, such as if the attackers seemed to know what they were looking for and how long it took them to undertake certain steps, e.g., to pivot to the next network segment and move on to further target machines Typical artifacts and hints that may aid the attribution process:

- **Hints on tactics, techniques and procedures (TTPs):** Log files, close monitoring of the network and affected hosts as well as forensic investigation results of malware used (e.g., file remnants) and exploited machines (e.g., bash history, left tools) might reveal important information on the modus operandi, e.g., which tools where used in which order to execute which actions? Based on that also the very important question whether insider knowledge was likely required, can be answered to a greater detail (Ligh et al. 2010).
- **Hints on organization and division of labor:** From the observed behavior (e.g., timing of actions) of the attackers revealed by log files, bash histories, recorded user sessions and process behavior monitoring, often behavior profiles can be extracted which allow some assumptions on the number of people involved in an operation as well as their patterns of life (working hours, shifts, timezone etc.) (Rid and Buchanan 2015). The application of shared scripts, reoccurring scanning methods and re-occurring tool parameters (e.g., in a reverse shell) can reveal information on how well a group is organized and their level of proficiency.
- **Hints on covert operations:** Investigations whether log files were modified to hide tracks (and how to detect this), or techniques used to distract monitoring mechanisms give important insights on the sophistication level of the attack and if the operation was meant to stay below the radar for an extended period (Gross 2011). Hiding tracks and avoiding collateral damage, which might lead to early detection, is resource-intensive and aids the construction of attacker profiles. Complex target verification (Yadav and Rao 2015) when malware becomes active is key to avoid collateral damage but is complicated to achieve. Often only secret services

have the capabilities and resources to carry this out on a high sophistication level.

Command & control artifacts

Taking permanent control over the target is an essential step for the attacker. However, taking control means communication with the outside world using any sort of obvious or covert channel. The most relevant questions of an investigator concerning this phase of an attack are centered on the actual C&C infrastructure used. Since effective C&C infrastructures are expensive to set up and maintain they are often re-used or rented. Furthermore, the applied communication and evasion techniques leave a quite unique "fingerprint" of the attacker. Typical artifacts and hints that may aid the attribution process:

- **DNS logs:** Logs provide vital insights into communication with external domains and allow correlation with threat intelligence about known malicious urls. Furthermore, passive DNS provide information on what domains where associated with which IP address at a certain time, thus allows to keep track of the usage of certain domains for malicious activities if they are investigated later.
- **Domain information:** The domain which is used to host C&C servers and often to "park" IP address pools to switch between C&C servers is connected to many further vital information particles, including registrar and payment information.
- **Hints on evasion techniques:** including covert channels and network steganography (Lubacz et al. 2014) provide insights into the attacker's capability level and may help to correlate breaches caused by the same groups.

Actions artifacts

In the last step, the attackers carry out the actions required to reach their goal, being system damage or data exfiltration. Once malicious actions have be identified, questions centered on what was the likely goal (exfiltration, manipulation, interruption, destruction), and did the attackers try to cover their tracks (i.e., did they care to maintain further access). In this phase, often attackers reveal information about themselves, e.g., by using a known drop zone for exfiltrated data or carrying out actions correlated to physical events (e.g., causing damage though a cyber attack as an answer to political developments). Typical artifacts and hints that may aid the attribution process:

- **Abnormal traffic and flows:** Unknown traffic and netflows are a rather obvious sign of malicious actions and can be gathered from firewalls, proxys and DNS servers (list of domain resolution requests). The target addresses of external machines may (indirectly) point to the intruders.
- **Log data from exploited machines:** If a machine is suspected of being exploited, log data can reveal vital insights into executed processes, associated users and accessed files; and thus, reveal important insights into applied TTPs (Kemmerer and Vigna 2002; Caltagirone et al. 2013).
- **Tools used and their configuration:** If tools were left on the target machine or their execution caused traces in log files and triggered events in monitoring solutions, e.g. the execution of prepared scripts, conclusions on the TTPs can be supported (Ligh et al. 2010).
- **External infrastructure used:** External infrastructures, such as cloud services used as drop zones for file exfiltration, DNS infrastructure and the like, may lead to actors, especially if the service providers are cooperative. At some point services must have been acquired and paid electronically.

Attribution and false flags

Attribution is about asking the right questions and coming up with plausible answers. It is very well comparable to playing a puzzle game – different pieces are discovered in a rather unordered fashion, but eventually they need to add up to a consistent picture.

Overview of the attribution process

The Cyber Attribution Model (CAM) (Pahi and Skopik 2019) consists of two main parts: cyber attack investigation (part I) and cyber threat actor profiling (part II). The attribution happens by matching these parts (see Fig. 2). Each part consists of technical and socio-political contextual indicators and the components of the CAM approach. The primary aim of the **cyber attack investigation** is to answer the questions, *Who is the victim and Why*, as well as *What has happened and How*. Answering these questions is guided by the components (i) victimology, (ii) infrastructure, (iii) capabilities and (iv) motivation. They help to discover TTPs, the modus operandi of a particular cyber attack and required capabilities – and possible false flags. The aim of the **cyber threat actor profiling** is developing profiles based on past attacks and find the matching profile to the findings from part I. The profiling helps finding answers to *Who could be the perpetrator*, *What infrastructure have they used* for the attack and W*hat capabilities and motivation might they have*. Cyber Threat Actor Profiling takes place either continuously or ad-hoc to support investigations. In that case part I (bottom-up) and part II (top-down) are running in parallel to find a match between the applied TTPs and possible perpetrator profiles.

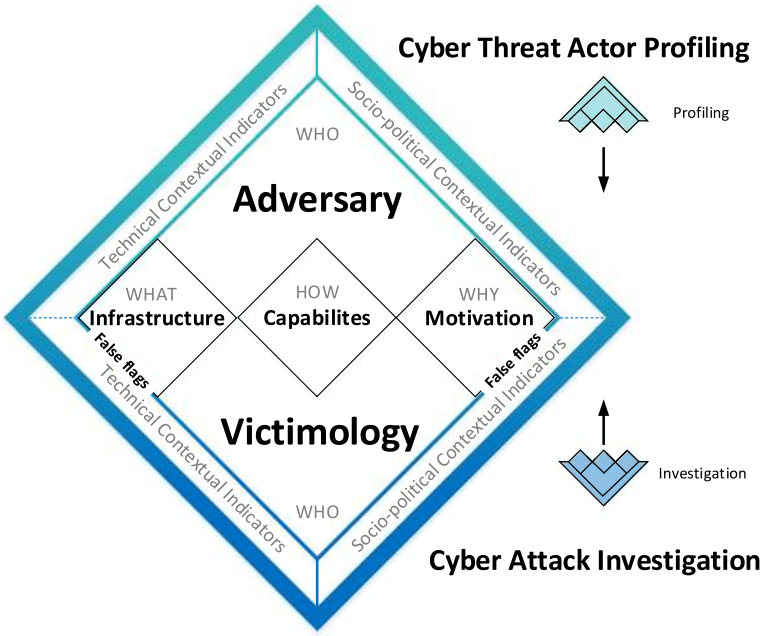


Fig. 2 The Cyber Attribution Model applied to detect inconsistencies in the attribution process

In both parts, technical and socio-political contextual indicators help to understand the evidences and to recognize complex correlations and possible false flag operations. The first step is often analysing the technical hard facts, aka applying digital forensics (Rid and Buchanan 2015). In this step, security specialists concentrate on the hard facts of already executed cyber attacks as an initial point. Various technical indicators of the attacks are analysed, such as applied malware, timestamps, strings, debug path, metadata, infrastructure and backend connection, tools, coding, language settings and pattern-of-life (Bartholomew and Guerrero-Saade 2016). The difficulty to manipulate or fake technical indicators greatly varies, depending on the infrastructure of the victim and perpetrator. We will focus on this aspect in greater detail later in this section.

The Socio-political contextual indicators cover the use of cyber tools for influencing the perception, opinion and behaviour of a target audience. False flag operations on this side belong to the categories of information war and influence cyber operations (Brangetto and Veenendaal 2016). Information has been manipulated for political purposes throughout the history of mankind, and the technological revolution opened new possibilities for state and non-state actors to use the cyberspace as a tool to shape the social and political mindset (Cohen and Bar'el 2017). There are numerous examples every day using a wide variety of forms of communication over the Internet and social media, such as influencing elections and spreading propaganda against or for political groups or ideologies, etc. However, processing these indicators is out of scope of this paper, which focuses almost exclusively on technical ones.

Cyber attack investigation and cyber threat actor profiling with CAM

In each cyber attack investigation, the starting point is the victim. Here, experts try to reconstruct the events that led to an incident. At this point victimology (Jaishankar 2011) comes into effect. Victimology in the cyber space is found in the literature primary in conjunction with cyber crime, especially cyber stalking. The term victimology stems from criminology and covers studying victims of crimes, the psychological effects of the crime. Professionals say, that there is no difference between a physical crime and a digital one (Halder and Jaishankar 2011). Just as an individual person has victimology-based characteristics, so do organizations. An organization's business interests, political action campaigns, vigilance level, protection abilities, and cyber risk tolerance are just some of the characteristics that can determine if an organization is more likely to be attacked, by whom, how, and why (Bullock 2018). The complementary part of the victimology is the threat actor profiling. Its aim is to analyse who is likely to commit a crime and what are the requirements for this. The victims of cyber attacks range from private businesses to nation states. Ukraine, France and the United States were affected by attacks during their elections, for instance.

The analysis of the technical contextual indicators and the victim's infrastructure help to reconstruct the operations, while the analysis of the Socio-political Contextual Indicators help to understand possible motives for the attack. After that, the required capabilities can better be pinpointed, such as the minimum requirements to execute the applied technical attacks and social engineering attacks or Influence Cyber Operations. At the end of the cyber attack investigation, the experts have collected all information about the victim and all technical and socio-political contextual indicators relevant for the incidents. Further results are Tactics, Techniques and Procedures (TTPs), i.e., the behavior or the modus operandi of the perpetrators, potential false flags on both side, and theories about possible opportunities and motivation. Deriving TTPs answers mainly what happened and how, helps to find the potential threat actors and to prevent similar attacks. Tactics have the highest abstraction level. It is the way an adversary chooses to carry out an attack, for instance, to use a malware to steal credit card credentials. Techniques are at a lower level of detail and procedures cover the related preparing processes and technological approaches for achieving intermediate results. An example would be sending targeted emails to potential victims with a malicious code attached. The procedure covers the organizational approach of the attack, for instance a special sequence of actions. This might be reconnaissance to identify potential individuals or creating an exploit to evade malware detection tools. To sum up, TTPs are used to describes an approach of threat actors, and finally also well suited to profiling threat actors.

The actual attribution takes place when findings from the investigation are matched to potential threat actor profiles. This requires to balance aspects of criminology, psychology and forensics (Turvey 2011) and studies mainly the motivation and methodology of the attackers. Profiling cyber threat actors is similar to profiling other fields. Since technology changes rapidly, IT security specialist must constantly keep up with the latest attack techniques (Long 2012). So, cyber threat actor profiling aims to create, update and manage threat actor profiles periodically. It is the complementary part to victimology and helps to better understand what type of threat actor the perpetrator could be. This action pinpoints the minimum required capabilities and observed TTPs. The analysts compare these results to known threat actor profiles. The analysis look for the matching applied infrastructures, tools and tactics. For instance, do the perpetrators have the required special knowledge for preparing the attack against rare industry components (e.g. Stuxnet), do they have the resources to develop their own toolkits and zero-day exploits or do they use already existing components. Furthermore, the analysis of the motives of the threat actors, such as political motives for hacktivist groups or ideological motives for certain hacker groups is part of actor profiling.

In case, the result from the cyber attack investigation and the potential threat actor profiles do not fit together, the analysts have to consider potential false flag actions. The CAM model distinguishes two types of false flags, one applied in technical context and one in socio-political context. There is a wide range of misdirecting actions. Therefore, careful attribution must have a particular focus on the consistency of the whole storyline. The presented CAM focuses mainly on targeted and sophisticated cyber attacks and covers additional social aspects and possible false flag operation for reliable attribution.

The technical dimension of attribution

From a technical view on the attribution process (Fig. 3) investigators try to collect as much (case-specific) artifacts as possible. For that purpose, they will find out which data sources are available in a first step. The MITRE Att&CK framework (MITRE 2019) provides a decent overview of technical artifacts collected in today's infrastructures, which might contain hints to the actual perpetrators. Additionally, further (mostly external) data

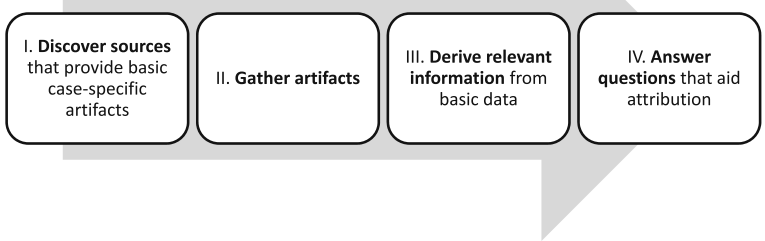


Fig. 3 A simplified linear view on the attribution process

sources, including threat intelligence feeds, social networks and news feeds need to be considered. This is a vast field and due to the wide range of available works out of scope of this paper. However, once potentially relevant sources have been identified, the investigators will gather the artifacts, derive higher-level information and raise major questions in course of the attribution process as described in the next few sections.

Gather artifacts

By systematically investigating and categorizing the artifacts identified in "Artifacts for the attribution process" section we came up with a structural view. First, all the artifacts that can be collected directly from technical infrastructures (Fig. 4), either internal ones of the victim or external ones, such as cloud providers, DNS or other service providers, especially e-Mail (if not hosted in-house), should be gathered. Second, artifacts may also be derived from further sources (Fig. 5), including the darknet, social networks, general news pages or cyber threat intelligence feeds. This list is not exhaustive but provides an impression why it is useful to take these sources into account during the analysis, e.g., for the estimation of the actor capabilities and/or motives.

Derive information

The gathered artifacts may be aggregated, interpreted and further processed to come up with the vital information for the attribution process, given in the first column of Table 1. However, not every piece of information is of the same value – a key criterion of using certain artifacts in the attribution process is their trustworthiness. We define this trustworthiness of an artifact as indirect proportional to the extent this artifact can be spoofed or tampered with. If it is easy for an attacker to manipulate an artifact its use in the attribution process is limited. Notice that rating the artifacts' trustworthiness is a very complex and ultimately infrastructure- and case-dependent metric. If false technical traces, such as spoofed log entries, can indeed influence the attribution does not only depend on the technical configuration, but also the skills of the forensic investigators to spot manipulation attempts.

In order to come up with a measure to rate the trustworthiness of the artifacts surveyed in "Artifacts for the attribution process" section, we conducted a structured survey applying the Delphi Method. The target group for the survey consist of fifteen IT security experts from a Forensics course at a SANS summit late 2019 with experience in forensic investigations, incident response and threat hunting. In a dedicated workshop, we explained the five APT cases referenced in "Related work" section and let then the experts individually decide how they would rate the reliability of an artifact in the given scenario, i.e., how much they would trust a particular type of information referenced in the first column of Table 1.

In particular the following questions were asked concerning technical artifacts from different sources to provide them some guidance for answering:

- Q1: How much effort (e.g., large team required, high amount of working hours) is it for an attacker to

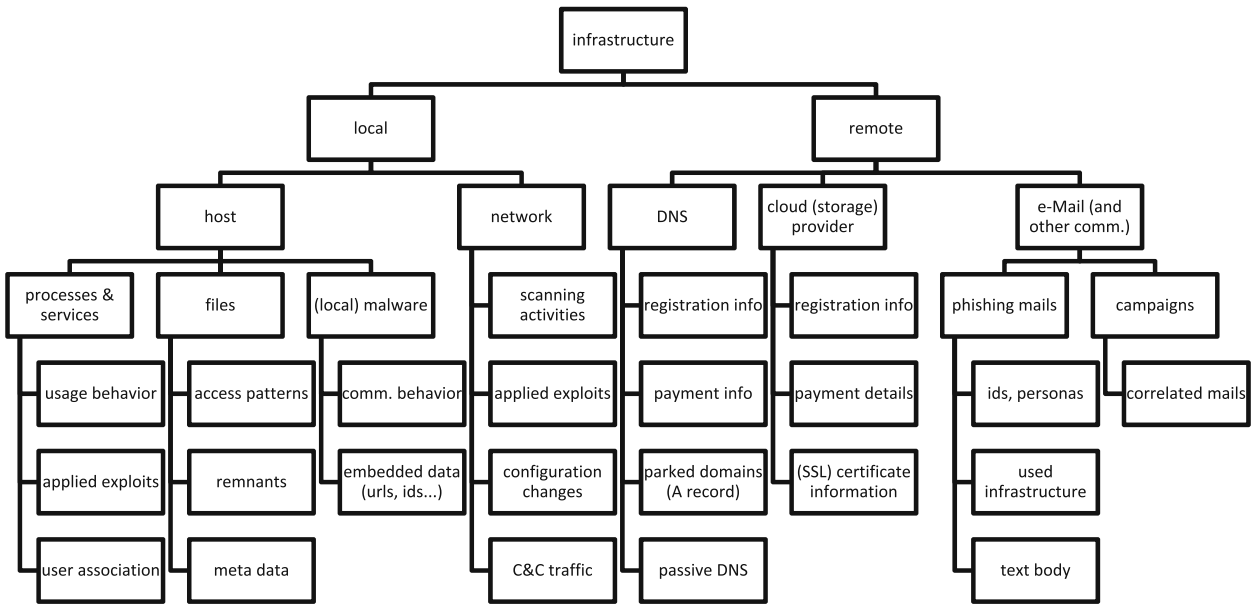


Fig. 4 Types of artifacts gathered from technical infrastructures that aid the attribution process

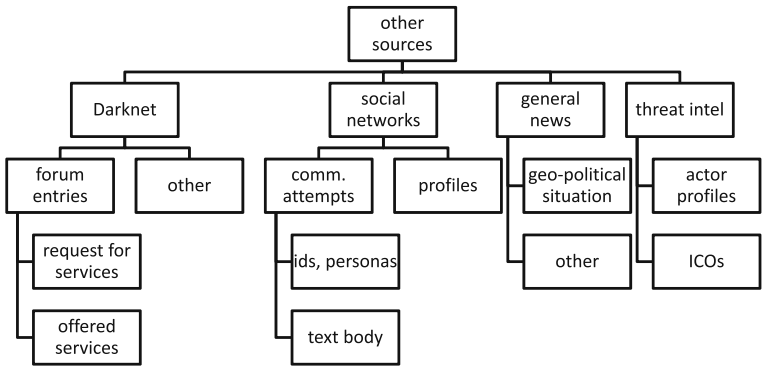


Fig. 5 Types of artifacts gathered from other sources (than technical infrastructures) that aid the attribution process

either spoof an artifact or change his/her actions to create different traces?

- Q2: How much special knowledge is required to manipulate related artifacts and leave only minor traces?
- Q3: How hard is it for the defenders to discover traces of manipulation or disguise?
- Q4: How unique and/or detailed are the potential traces?

Table 1 Classes of information, derived from artifacts from various sources and an estimation of the trustworthiness of this information for attribution

Type of information for attribution	Sources of artifacts	trustworthiness
(T1) General TTPs (typical modus operandi) in a stepwise attack. This includes pattern-of life, focus on certain services/applications, usage of zero day exploits etc. For instance, one group may be proficient in developing browser exploits and deploying them in watering hole attacks; another one more focused on social engineering attacks.	API monitoring, application logs, authentication logs, DLL monitoring, DNS records, e-Mail gateway, file monitoring, HIDS, kernel driver, netflows, network device logs, NIDS, packet captures, power shell logs, process command line parameters, process monitoring, SSL/TLS inspection, Web logs, Windows registry.	4.2 (0.7)
(T2) Software tools frequently used (related to TTPs); Notice this is a highly controversial topic. On the one side attacker groups tend to reuse tools, which they know well. So, from the set of used tools and their combination one may be able to characterize attackers. On the other side, tools are also picked depending on the target infrastructure in order to reach a specific goal and thus may look differently for the same group but for different targets.	API monitoring, application logs, binary file meta data, disk forensics, file monitoring, HIDS, kernel driver, process monitoring, Web logs, Windows registry.	3.1 (0.5)
(T3) Phishing attempts in form of e-mails, social networking, messengers, and therein certain spellings, usage of words, grammar mistakes, writing styles etc.	e-Mail gateway, social networking sites/crawlers, Web proxy, malware analysis.	2.1 (0.9)
(T4) Identities, pseudonyms and Personas, potentially reused from previous attacks.	e-Mail gateway, social networking sites, Web proxy, numerous messenger services, payment/billing information from providers.	3.2 (1.8)
(T5) Cloud services and C2 infrastructure used, including re-use of domains, usage of certain botnets.	authentication logs, DNS records, server logs, payment information from cloud provider.	4.6 (0.4)
(T6) DNS patterns, such as registration information, parked domains (A records), which are quickly changed during an attack; passive DNS data (Bilge et al. 2011).	DNS records, payment information from cloud provider, passive DNS service.	4.4 (0.5)
(T7) Local Malware and their properties, such as compiler language, programming language, compile time, libraries used, keyboard layout, ...	malware reverse engineering.	2.3 (1.3)
(T8) Traces in the darknet consistent with technical artifacts, e.g., attempts in forums to acquire zero day exploits, hire a hacker, rent a service in the planning phase; or even the attempt to sell confidential information after a successful breach. typical traces: usernames, bitcoin wallets etc.	forum/board entries gathered through "spiders" and mining (Nunes et al. 2016).	1.2 (2.1)
(T9) Encounters in the real world, e.g., blackmailing of employees, political statements, verbal threats, baiting, physical security breaches, ...	monitoring of news feeds and analysis for keywords.	3.3 (1.2)

- Q5: How closely related are other artifacts (important to draw a consistent picture)?

Each of these five questions was answered by each expert for artifacts from the nine classes (T1-T9) of sources given in column 2 of Table 1. The average value of each expert was then calculated and the scores of all experts fed back in a second round, where everyone had the chance to revise his/her answers.

The final score, reflecting how easily an artifact can be consistently faked, was then calculated as weighted average where the experts had the opportunity to state their certainty about their estimation on a scale from one to five. In other words, if an expert was not sure about his/her estimation, s/he had less influence on the final score. This final score represents the estimated trustworthiness for the different classes of information is given in the third column from 1 (*low* trustworthiness) to 5 (*high* trustworthiness) – in context of the given five cases, which represent a good mix of different APT cases. Besides the average rating, we also provide the standard deviation to show how much experts agreed on the final evaluation. The standard deviation is given in parenthesis.

In short, besides the usage of external technical infrastructure (cloud services, DNS etc.), actual TTPs, i.e., the tactics and applied techniques by attackers are hardest to fake as these represent how a group operates. At the other end of the scale, traces in the darknet can easily be spoofed, e.g., by impersonating other actors. Somewhat in the middle are the actually applied tools, which may on the one side represent a groups 'IPRs', on the other side be acquired, rented and adopted, and therefore a potentially unreliable indicator. The highest standard deviation, i.e., most differing views from experts was related to encounters in the real world. While some argued that a correlation of cyber attacks with real world riots and terrorist attacks provides an opportunity for plausibility checks, others pointed out that this is also a great opportunity for attackers to place false flags.

Answer questions

Based on a thorough evaluation of technical traces, attribution aims at answering the questions (among others) in Table 2 to better understand the attacker's perspective. In this process, the analysis of the artifacts discussed before, provide vital answers on questions concerning the victim's (and any third party) *infrastructure*, the actor's *capabilities* and their specific *motivation* to carry out the investigated attack.

Many single properties of a cyber attack can be spoofed, faked and disguised; for instance, IP addresses by using (chains of) proxy servers or the TOR network, people can be impersonated, as well as language settings, agent strings, and false hints in code placed. However, it is quite complicated to get all these tiny pieces consistently right. Eventually, careful attribution must have a particular focus on the consistency of the whole storyline. If a single factor looks odd or does not fit to the obvious story – something is in fact odd.

Table 2 Characterizing threat actors based on gained insights

Questions regarding the relevant INFRASTRUCTURE:

- Was the victim specifically selected or likely hit by chance? (e.g., Was there a mechanism for target validation?)
- Where other target infrastructures attacked in parallel?
- Was insider knowledge likely required to break in effectively?
- What external infrastructure (Cloud, DNS etc.) was utilized to carry out the attack?

Questions regarding the CAPABILITIES of a threat actor:

- What special skills were required in order to build the payload?
- How rare are these skills and who is known to have them?
- What budget and time resources were likely required?
- Who has the facilities and access to certain components (in case of CPS) to re-build the target environment and test exploits accordingly? (if applicable)
- Where there any beginner's mistakes made?
- Where certain actions sloppy compared to other steps in the whole process?
- Was the operation successful in general?

Questions regarding the MOTIVATION of a threat actor:

- Who has a clear benefit to breach into the target organization?
- Who or what has most harm, also considering side- and long-term effects?
- Who or what was damaged most and which impact can be predicted for the future?
- Can any current political developments be associated to this attack?

Illustrative application of CAM to identify false flags

After we let experts rate the trustworthiness of different types of information in the attribution process based on five common APT cases (cf. "Related work" section), we are going to prove the applicability of our findings in a sixth case. In this case we take a closer look into relevant artifacts and their relation to the suggested categories of artifacts **T1-T9**, as well as their degree of trustworthiness for attribution in Table 1. The selected scenario, the well-documented TV5Monde hack (InfosecPartners 2016), serves as a basis for a short presentation of the application of the Cyber Attribution Model (CAM) (Pahi and Skopik 2019) (cf. Fig. 2) and how we identify and tackle potential false flags. The victim organization and the French intelligence services have shared their experiences about this incident. The application of CAM starts with Cyber Attack Investigation, especially with victimology. The victim is TV5Monde, a French television network claiming to

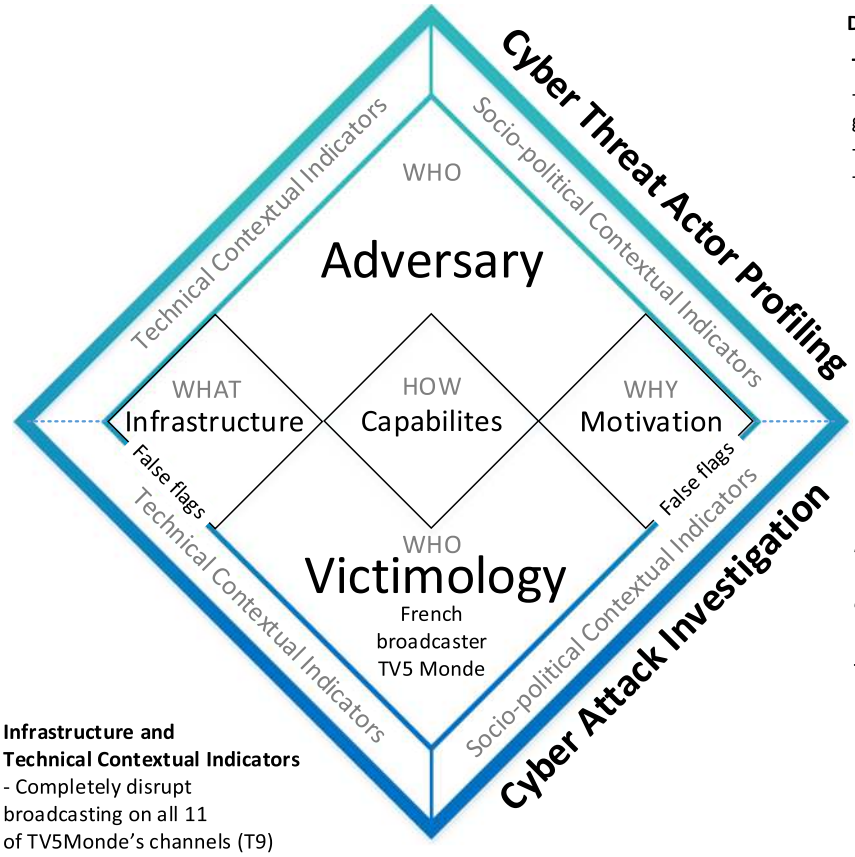
be one of the top three most available global television networks internationally. TV5Monde was a victim of a cyber attack, which caused service disruptions for hours in April 2015.

The circumstances of this incident can be understood by analysing socio-political contextual indicators and motivations. At the time of the attack, France was still in shock from terrorist attacks (7th January 2015) on the editors of Charlie Hebdo, a French satirical weekly magazine. The French national agency for the security of information systems reported more than 1,500 cyber attacks against small companies' websites in the wake of the attack on the Charlie Hebdo office in Paris (Maurice 2015). Further, the attack on TV5Monde was followed by a series of terrorist attacks, such as mass shooting and suicide bombing in Paris on 13th November 2015. In parallel to the TV5Monde attack, pro-IS messages appeared on the TV station's Twitter and Facebook accounts. Several of the posts included messages against the United States and France, as well as threats issued to families of French soldiers. Furthermore, copies of French soldiers' IDs and passports were published. All these socio-political encounters in the real world (**T9**) are essential to get a holistic picture and to account for all potential perpetrators.

Applying the CAM approach, the cyber attack investigation continues with the analysis of technical evidences and the victim infrastructure. The technical contextual indicators cover the information discovered by the real incident response team involved in the investigation and reconstruct what has happened. This part of the investigation includes the following types of information for the attribution for **T1**, **T2** and **T6**. The attackers got their initial access on 23rd January 2015 and took over a server used by the broadcasting company. One of the TV5Monde multimedia servers had its RDP port exposed to the Internet and was configured with a default username and password. Since the server was not connected to the internal network, the attackers continued the reconnaissance (providing artifacts mainly of types **T1** and **T2**). They returned later, using a compromised third-party account that allowed them to connect to the TV5Monde VPN on 6th February (related to **T4**). After that, attackers began scanning internal machines connected to an infected endpoint and identified two internal Windows systems, that were used to manage cameras. Afterward, the attackers used one of these compromised systems to create a new administrator-level user in the Active Directory (AD) called 'LocalAdministator' on 11th February 2015 (InfosecPartners 2016). The first major clue was, that all AD administrator names had French descriptions except for one. During the reconnaissance phase of the attack (16th February - 25th March 2015) the attackers mapped the network's IT services in the victim's infrastructure and collected related information, including information from the IT department's internal wiki, which provided details on how logins and passwords were handled (Schwarzt 2017). After that, the attackers compromised another administrator machine with a remote access control software (RAT), that was used for the sabotage (related to **T7**). At 19:57 on 8th April, the attackers performed their first damaging operation by reconfiguring all the IP settings of the media in a faulty manner. This misconfiguration was only enabled, when the technical teams rebooted their machines. At 20:58, the online presence was affected through hacked social media accounts (YouTube, Facebook, Twitter), and the website of TV5Monde was defaced. At 21:45, the attackers run commands via TACACS logs, that erased switch and router firmware, resulting in black screens for viewers, except for one new channel that was launched on the same day.

Related to applied malware (**T7**), the investigations definitely showed the application of Sednit (aka Sofacy) malware (ESET 2016), associated with the ongoing Pawn Storm campaign (TrendMicro 2015). Operation Pawn Storm is an active economic and political cyber-espionage operation that targets a wide range of high-profile entities, from government institutions to media personalities, referred to as APT 28. It is however remarkable that there is no direct connection between the Pawn Storm campaign and the TV5Monde attack (TrendMicro 2016).

These findings (see overview in Fig. 6) shape possible hacker profiles and so lead to the cyber threat actor profiling. The CAM verifies now the results of the cyber investigation (part I) by asking questions regarding the three main components of the cyber threat actor profiling (part II), namely the perpetrators' infrastructure, capabilities and motivation (cf. Table 2). To sum up the required capabilities in this case, the attackers had to have access to toolsets applied in the Pawn Storm espionage operation and to copies of French soldiers' IDs. Further, they needed the capability and resources (1) to execute an at least 3-months lasting stealthy attack with a deep reconnaissance phase, (2) to have solid knowledge about the victim, by attacking their subcontractors and (3) to carry out a complex network compromise and website-, social media defacement. Related to **T7**, the cyber threat actor profiling part supports the analysis on potential threat actors that fulfil these requirements. The actual perpetrator needed the required infrastructure (e.g. Sednit malware), the capabilities (e.g. well-organized group with deep technical knowledge) and the motivation (e.g. ideological motivated or red herring). The result would be verified asking special questions based on the investigation, such as which threat actors have access to the Sednit malware (related to **T8**).



Database of Threat Actor Profiles

Threat Actor Profile: APT28
-Target:wide range of military and government targets in Europe and America
- Motivation: political, espionage
- Alias: Sofacy, Fancy Bear

Threat Actor Profile: CyberCaliphate
- Target: against institutions viewed opposing values and political ideologies
- Motivation: ideological
- Alias: United Cyber Caliphate

Possible False Flag *(Motivation)*
(T4) CyberCaliphate initially claimed resposibility
(T4) Pro ISIS messages on the social media

Possible False Flags *(Infrastructure)*
(T2, T7) Application of Sednit. Malware is associated with Pawn Storm campaign
(T5, T6) CyberCaliphate website hosted on the same IP block as other known APT28 infrastructure

Infrastructure and Technical Contextual Indicators
- Completely disrupt broadcasting on all 11 of TV5Monde's channels (T9)
- Take control of TV5Monde's website and social media accounts
- Replace content on the website with pro-ISIS statements (T4)
- Phishing and social engineering against TV5 Monde (T3)
- Breaching months ago via VPN through one of TV5's subcontractor (T5, T6)
- Creating a new admin-level user in AD (T4)
- Comprimising and infecting another admin machine (T1)
- Compromising signal transmission server and gateway connections (T9)

Capabilities
- Access to the credentials from the Pawn Storm incident (could be T8)
- Capability to carry out a highly visible website, social media and TV network compromise
- Having copies of French soldiers' IDs
- Having staff for a well-organized at least 3-months lasting stealthy attack (T1)
- Knowing TV5 Monde's subcontractors

Motivation and Socio-political Contextual Indicators
- France was still in shock from terrorist attacks
- French national agency reported more than 1,500 cyber attacks perpetrated by groups claiming to act "in defense of Muslims" or for the Palestinian cause
- Series of terrorist attacks, such as mass shooting and suicide bombing, in Paris 13th November 2015

Fig. 6 CAM applied to the TV5Monde APT case

Interpretation of the results of investigation

In case of the TV5Monde hack, there are three possible theories available with two known potential threat actor profiles. One potential threat actor is the CyberCaliphate. This is a hacker group targeting institutions with opposing ideologies. Another one is the APT28 group targeting military and governmental facilities in Europe and America.

1 The first theory is that TV5Monde was the victim of two entirely unrelated incidents. TV5Monde was victim of Pawn Storm operation and a separate hacktivist attack.
2 The second theory is that the Pawn Storm group gave information, which was relevant for the attack, to a third party (to an unknown perpetrator), directly or indirectly connected to Islamic hacktivists. While possible, this would seem highly unlikely as Pawn Storm actively targeted Chechen separatists and Islamic extremists in former Yugoslavia in the past.

3 Third, the Pawn Storm group carried out the attack and used it as a false flag operation to lay the blame on Islamic extremists (TrendMicro 2015).

It has become the consensus view among Western intelligence services that the CyberCaliphate and the TV5Monde hack were Russian intelligence's false flag operations. The hypothesis is that the Russian intelligence agencies go to cyber war against the West under an IS cloak (Obervser Schindler 2016). In that case, the attribution could connect the infrastructure, capabilities and motivation with the matching threat actor. The possible underlying motive, to test out cyber capabilities or assess Western intelligence agencies' ability to spot such misdirection, remains unknown (Corera 2016).

Assessment of false flags using the trustworthiness metric

At the time when the attack took place, the CyberCaliphate group initially claimed responsibility for the attack. This fact points to one of the most controversial types of information for attribution, namely to identities (**T4**). This information type has the highest standard deviation in our survey (see Table 1) – for a good reason. In our case, only the claim of responsibility of the CyberCaliphate on their websites and on the hijacked social media accounts links the attack to the hacktivists. The perpetrators used stolen social media accounts and posted in the name of the CyberCaliphate. Such identity thefts on social media platforms are quite popular and easy to carry out today (Irshad and Soomro 2018). Therefore, the apparently main trace pointing to the hacktivist is not trustworthy. The investigation has to focus on other types of information. According to our survey, the TTPs are the hardest traces to fake (**T1**). The experts of FireEye found out that there are a number of similar TTPs used by APT28 and the perpetrators in this incident. The CyberCaliphate website, where they posted the data on the TV5Monde hack, was hosted on an IP block which is the same IP block as other known APT28 infrastructure, and used the same server that APT28 used in the past (Paganini 2015). It would mean, that this attack was the work of undisciplined Pawn Storm actors. Although the Pawn Storm actors normally work in a professional way, there have been a few other incidents where some Pawn Storm actors showed a lack of discipline (TrendMicro 2016). Further investigations on the TTPs revealed links to the Russian hacking group APT28 despite the CyberCaliphate's confession (Council on Foreign Relations 2018). Security experts state, that the attackers' have their strengths not just in their choice of tools or in their experience, but in their capability to execute this long-time reconnaissance and the synchronized attacks (MacFarquhar 2016).

In this case, the actual TTPs are the most reliable information for the further analysis – they also received the highest trustworthiness score in our survey. On the one hand, the perpetrator used the same IP range as APT28 in a former attack. This type of information points to re-using domains, as indicated by **T5** and **T6**. These traces are hard to fake and were presumably visible only for APT28 and the investigators of their attacks. One the other hand, discipline and the level of professionalism belong also to TTPs. A high level of professionalism is also hard to fake, therefore this fact has a high trustworthiness score. In contrast, low level of discipline is easy to imitate. The easily falsified use of fake identities (**T4**) was a hint that the attack could be a false flag operation. The application of the Sednit malware is also not such a strong indicator (accordingly the evaluated trustworthiness score of **T7** is rather low in Table 1), therefore not a major link to the perpetrator. Eventually the relevant types of information used for a reliable attribution in this case were related to the utilization of external infrastructure (**T5** with a trustworthy score of 4.6 and **T6** with a score of 4.4) and the general TTP (**T1** with a score of 4.2) – all artifacts associated with accordingly high trustworthiness scores, which validates the opinions of the experts in our survey.

Summary of findings regarding false flag activities

Identifying false flags is not a simple task. Not only try perpetrators to disguise their activities or let a third party take the blame for malicious actions (Rid and Buchanan 2015), but false flags can also be created unintentionally. From analyzing recent APTs (refer to "Related work" section), we conclude that false flags exist for one or several of the following reasons:

- Exploits use recycled code/variants from previous attacks that worked in the past and became public (Brown et al. 2015).
- Exploits are developed to mimic the behavior and complexity of known, attributed malware.
- Exploits and malware are bought rather than developed (Miller 2007; Algarni and Malaiya 2014).
- An attack is rented as a service (Santanna et al. 2015).
- Malware connects to a known C&C infrastructure, although it was not designed by the server operators (Stone-Gross et al. 2009).
- A C&C server uses infrastructures of a third party not related to the attackers, e.g., exploited Web servers in another country.
- A breach is socially engineered to misguide investigations towards other operators (Kijewski et al. 2016).
- The execution of further malicious actions hide the actual intent to mislead investigators.

Conclusion and future work

In this paper we surveyed which artifacts after a security incident, i.e., a breach and/or intrusion, allow to learn more about the attacker. We highlighted which artifacts exist, how they contribute to the attribution process and discussed their reliability. In general, the investigation of previous APTs along with expert feedback draw the following picture: Artifacts related to the application of technologies (tools, malware, exploits) may be easier to spoof compared to the general modus operandi (TTPs) on a higher level. Furthermore, traces within the victim's infrastructure (besides the application of tools also the type and modus of the lateral movement) are easier to fake than traces left outside the victim's organization. The latter means the setting up a C&C infrastructure is not only cost-intensive but also needs connections to real persons (acquiring domains, renting cloud services etc.). However, information, such as past DNS-IP associations and linked registrar information is much harder to acquire for the investigators; often only a nation state – if at all – is able to gather this information, no private investigator.

Not only spoofing and faking is the problem, but also the frequent reuse and adoption of TTPs of one attacker group by another one. Then, attribution becomes extremely though. Another dilemma is the exploitation of third party infrastructures for malicious activities without the victim's knowledge, e.g., Mail servers exploited for spamming activities by a third party.

Many cyber attackers use advanced programming techniques to create the right tool for the right operation, but sometimes they overlook small details and make very basic mistakes. To make attribution more accurate in the future, we need to focus on the actors and the context, not isolated technical information only. Behavior profiles rather than simple IOCs are going to produce higher fidelity results.

Acknowledgements
Not applicable.

Authors' contributions
Jointly written by F. Skopik and T. Pahi. The author(s) read and approved the final manuscript.

Funding
This work was partly funded by the Austrian security-research programme FORTE and by the Austrian Ministry for Transport, Innovation and Technology (BMvit) through the FFG project CADSP (873425).

Availability of data and materials
Not applicable.

Competing interests
None

Received: 8 October 2019 Accepted: 20 February 2020
Published online: 20 March 2020

References

2016 Public-Private Analytic Exchange Program Team (2016) Cyber Attribution Using Unclassified Data. https://www.dni.gov/files/PE/Documents/Cyber-Attribution.pdf. Accessed 25 Feb 2020
Afroz S, Brennan M, Greenstadt R (2012) Detecting hoaxes, frauds, and deception in writing style online. In: 2012 IEEE Symposium on Security and Privacy. IEEE. pp 461–475
Algarni A, Malaiya Y (2014) Software vulnerability markets: Discoverers and buyers. Int J Comput Inf Sci Eng 8(3):71–81
Alperovitch D, et al. (2011) Revealed: Operation Shady RAT, vol. 3. McAfee, Santa Clara
Bartholomew B, Guerrero-Saade JA (2016) Wave your false flags! deception tactics muddying attribution in targeted attacks. In: Virus Bulletin Conference. pp 1–9
Bartlett G, Heidemann J, Papadopoulos C (2007) Understanding passive and active service discovery. In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. ACM. pp 57–70
Beatty M (2019) The current and future threat of steganography in malware command and control. PhD thesis. Utica College
Bilge L, Dumitraş T (2012) Before we knew it: an empirical study of zero-day attacks in the real world. In: Proceedings of the 2012 ACM Conference on Computer and Communications Security. ACM. pp 833–844
Bilge L, Kirda E, Kruegel C, Balduzzi M (2011) Exposure: Finding malicious domains using passive dns analysis. In: Ndss. pp 1–17
Bowen BM, Hershkop S, Keromytis AD, Stolfo SJ (2009) Baiting inside attackers using decoy documents. In: International Conference on Security and Privacy in Communication Systems. Springer. pp 51–70
Brangetto P, Veenendaal MA (2016) Influence cyber operations: The use of cyberattacks in support of influence operations. In: 2016 8th International Conference on Cyber Conflict (CyCon). IEEE. pp 113–126
Brenner SW (2006) At light speed: Attribution and response to cybercrime/terrorism/warfare. J Crim L Criminol 97:379
Brown S, Gommers J, Serrano O (2015) From cyber security information sharing to threat management. In: Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security. ACM. pp 43–49
Bullock C (2018) Don't Forget Victimology as a Cybersecurity Strategy. Secureworks Inc. https://www.secureworks.com/blog/dont-forget-victimology-as-a-cybersecurity-strategy. Accessed 25 Feb 2020
Caltagirone S, Pendergast A, Betz C (2013) The diamond model of intrusion analysis. Technical report. Center For Cyber Intelligence Analysis and Threat Research, Hanover Md
Cherepanov A (2018) GreyEnergy-A successor to BlackEnergy. https://www.welivesecurity.com/wp-content/uploads/2018/10/ESET_GreyEnergy.pdf. Accessed 25 Feb 2020
Chiesa R, Ducci S, Ciappi S (2008) Profiling Hackers: the Science of Criminal Profiling as Applied to the World of Hacking. Auerbach Publications, Boca Raton
Cohen D, Bar'el D (2017) The use of cyberwarfare in influence operations. In: October, Tel Aviv: Yuval Ne'eman Workshop for Science, Technology and Security. pp 1–58
Corera G (2016) How france's tv5 was almost destroyed by 'russian hackers'. BBC News
Cova M, Kruegel C, Vigna G (2010) Detection and analysis of drive-by-download attacks and malicious javascript code. In: Proceedings of the 19th International Conference on World Wide Web. ACM. pp 281–290
Council on Foreign Relations (2018) Compromise of TV5 Monde. https://www.cfr.org/interactive/cyber-operations/compromise-tv5-monde. Accessed 25 Feb 2020
ESET (2016) En Route with Sednit. http://www.welivesecurity.com/wp-content/uploads/2016/10/eset-sednit-full.pdf. Accessed 25 Feb 2020
Falliere N, Murchu LO, Chien E (2011) W32. stuxnet dossier. White Pap Symantec Corp Secur Response 5(6):29
Foster I, Prudhomme A, Koscher K, Savage S (2015) Fast and vulnerable: a story of telematic failures. In: 9th USENIX Workshop on Offensive Technologies (WOOT 15). USENIX Association, Washington, D.C.
Galperin E, Marquis-Boire M (2012) Pro-syrian government hackers target activists with fake anti-hacking tool. Electron Front Found. https://www.eff.org/deeplinks/2012/08/syrian-malware-post. Accessed 25 Feb 2020
Goodman W (2010) Cyber deterrence: Tougher in theory than in practice? Technical report. Washington DC Committee on Armed Services, Senate (United States)

Gross MJ (2011) A declaration of cyber-war. Vanity Fair. https://www2.cs.duke.edu/courses/common/compsci092/papers/cyberwar/stuxnet.pdf. Accessed 25 Feb 2020
Haagman D, Ghavalas B (2005) Trojan defence: A forensic view. Digit Investig 2(1):23–30
Hadnagy C (2010) Social Engineering: The Art of Human Hacking. Wiley, New York City
Halder D, Jaishankar K (2011) Cyber gender harassment and secondary victimization: A comparative analysis of the united states, the uk, and india. Vict Offenders 6(4):386–398
Harrington SL (2014) Cyber security active defense: Playing with fire or sound risk management. Richmond J Law Technol 20(4):12
Harris R (2006) Arriving at an anti-forensics consensus: Examining how to define and control the anti-forensics problem. Digit Investig 3:44–49
Hulnick AS (2010) The dilemma of open sources intelligence: Is osint really intelligence? In: The Oxford Handbook of National Security Intelligence. Oxford University Press
Hunker J, Hutchinson B, Margulies J (2008) Role and challenges for sufficient cyber-attack attribution. Inst Inf Infrastruct Prot: 5–10
Hutchins EM, Cloppert MJ, Amin RM (2011) Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Lead Issues Inf Warf Secur Res 1(1):80
InfosecPartners (2016) Autopsy of Cyber Attack on TV5Monde. https://www.infosecpartners.com/newsroom/2016/10/10/autopsy-cyber-attack-tv5monde/. Accessed 25 Feb 2020
Irshad S, Soomro TR (2018) Identity theft and social media. Int J Comput Sci Netw Secur 18(1):43–55
Jaishankar K (2011) Cyber Criminology: Exploring Internet Crimes and Criminal Behavior. CRC Press, Boca Raton
Jarvis L, Macdonald S (2015) What is cyberterrorism? findings from a survey of researchers. Terrorism Polit Violence 27(4):657–678
Kang H, Kim H, Lee H, Lee S-j (2017) Study on collecting server information through banner grabbing. J Korea Inst Inf Secur Cryptol 27(6):1317–1330
Kaushik AK, Pilli ES, Joshi R (2010) Network forensic system for port scanning attack. In: 2010 IEEE 2nd International Advance Computing Conference (IACC). IEEE. pp 310–315
Kearns EM, Conlon B, Young JK (2014) Lying about terrorism. Stud Confl Terrorism 37(5):422–439
Kemmerer RA, Vigna G (2002) Intrusion detection: a brief history and overview. Computer 35(4):27–30
Kijewski P, Jaroszewski P, Urbanowicz JA, Armin J (2016) The never-ending game of cyberattack attribution. In: Combatting Cybercrime and Cyberterrorism. Springer, International. pp 175–192
Krombholz K, Hobel H, Huber M, Weippl E (2015) Advanced social engineering attacks. J Inf Secur Appl 22:113–122
Kumar V, Srivastava J, Lazarevic A (2006) Managing Cyber Threats: Issues, Approaches, and Challenges, vol. 5. Springer, New York
Lemay A, Calvet J, Menet F, Fernandez JM (2018) Survey of publicly available reports on advanced persistent threat actors. Comput Secur 72:26–59
Li F (2014) Apt attribution and dns profiling. Black Hat
Li C, Jiang W, Zou X (2009) Botnet: Survey and case study. In: 2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC). IEEE. pp 1184–1187
Liao K, Zhao Z, Doupé A, Ahn G-J (2016) Behind closed doors: measurement and analysis of cryptolocker ransoms in bitcoin. In: 2016 APWG Symposium on Electronic Crime Research (eCrime). IEEE. pp 1–13
Ligh M, Adair S, Hartstein B, Richard M (2010) Malware Analyst's Cookbook and DVD: Tools and Techniques for Fighting Malicious Code. Wiley Publishing, New York City
Long LA (2012) Profiling hackers. SANS Institute Reading Room: pp 1–22. https://www.sans.org/reading-room/whitepapers/hackers/profiling-hackers-33864. Accessed 25 Feb 2020
Lubacz J, Mazurczyk W, Szczypiorski K (2014) Principles and overview of network steganography. IEEE Commun Mag 52(5):225–229
Lyon GF (2009) Nmap Network Scanning: The Official Nmap Project Guide to Network Discovery and Security Scanning. Insecure, Sunnyvale
MacFarquhar N (2016) A powerful russian weapon: The spread of false stories. N Y Times 28:2016
Maurice E (2015) Cyber attack on French TV finds EU unprepared. https://euobserver.com/news/128285. Accessed 25 Feb 2020
Miller C (2007) The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales. In: In Sixth Workshop on the Economics of Information Security
MITRE (2019) ATT&CK Matrix for Enterprise. https://attack.mitre.org/. Accessed 25 Feb 2020
Morgan R, Kelly D (2019) A novel perspective on cyber attribution. In: International Conference on Cyber Warfare and Security. Academic Conferences International Limited. p 609
Nunes E, Diab A, Gunn A, Marin E, Mishra V, Paliath V, Robertson J, Shakarian J, Thart A, Shakarian P (2016) Darknet and deepnet mining for proactive cybersecurity threat intelligence. In: 2016 IEEE Conference on Intelligence and Security Informatics (ISI). IEEE. pp 7–12
Obervser Schindler JR (2016) False Flags:The Kremlin's Hidden Cyber Hand. https://observer.com/2016/06/false-flags-the-kremlins-hidden-cyber-hand/. Accessed 25 Feb 2020
Paganini P (2015) FireEye Claims Russian APT28 Hacked France's TV5Monde Channel. Secur Aff. https://securityaffairs.co/wordpress/37710/hacking/apt28-hacked-tv5monde.html. Accessed 25 Feb 2020
Pahi T, Skopik F (2019) Cyber attribution 2.0: Capture the false flag. In: ECCWS 2019 18th European Conference on Cyber Warfare and Security. Academic Conferences and publishing limited. p 338
Peterson D (2013) Offensive cyber weapons: construction, development, and employment. J Strat Stud 36(1):120–124
Philbin MJ (2013) Cyber deterrence: An old concept in a new domain. Technical report. Army War College, Carlisle Barracks PA
Pihelgas M (2015) Mitigating risks arising from false-flag and no-flag cyber attacks. CCD COE, NATO, Tallinn
Rid T, Buchanan B (2015) Attributing cyber attacks. J Strateg Stud 38(1-2):4–37
Santanna JJ, van Rijswijk-Deij R, Hofstede R, Sperotto A, Wierbosch M, Granville LZ, Pras A (2015) Booters—an analysis of ddos-as-a-service attacks. In: 2015 IFIP/IEEE International Symposium on Integrated Network Management (IM). IEEE. pp 243–251
Schwarzt MJ (2017) French Officials details "Fancy Bear" hack on TV5Monde. https://www.bankinfosecurity.com/french-officials-detail-fancy-bear-hack-tv5monde-a-9983. Accessed 25 Feb 2020
Skopik F, Settanni G, Fiedler R (2016) A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. Comput Secur 60:154–176
Stone-Gross B, Cova M, Cavallaro L, Gilbert B, Szydlowski M, Kemmerer R, Kruegel C, Vigna G (2009) Your botnet is my botnet: analysis of a botnet takeover. In: Proceedings of the 16th ACM Conference on Computer and Communications Security. ACM. pp 635–647
Strom BE, Battaglia JA, Kemmerer MS, Kupersanin W, Miller DP, Wampler C, Whitley SM, Wolf RD (2017) Finding cyber threats with att&ck™-based analytics. Technical report. MITRE Technical Report MTR170202. The MITRE Corporation
Tankard C (2011) Advanced persistent threats and how to monitor and deter them. Netw Secur 2011(8):16–19
Tran D (2018) The law of attribution: Rules for attribution the source of a cyber-attack. Yale JL Tech 20:376
TrendMicro (2015) TV5 Monde, Russia and the CyberCaliphate. https://blog.trendmicro.com/tv5-monde-russia-and-the-cybercaliphate/. Accessed 25 Feb 2020
TrendMicro (2016) Operation Pawn Storm: Fast Facts and the Latest Developments. https://blog.trendmicro.com/tv5-monde-russia-and-the-cybercaliphate/. Accessed 25 Feb 2020
Tsagourias N (2012) Cyber attacks, self-defence and the problem of attribution. J Confl Secur Law 17(2):229–244
Tsagourias N, Farrell MD (2018) Cyber attribution: technical and legal approaches and challenges. draft article. https://sites.tufts.edu/cilg/files/2018/09/attributiondraftsm.pdf. Accessed 25 Feb 2020
Tuli P, Sahu P (2013) System monitoring and security using keylogger. Int J Comput Sci Mob Comput 2(3):106–111
Turvey BE (2011) Criminal Profiling: An Introduction to Behavioral Evidence Analysis. Academic press, USA
Vogt P, Nentwich F, Jovanovic N, Kirda E, Kruegel C, Vigna G (2007) Cross site scripting prevention with dynamic data tainting and static analysis. In: NDSS, vol. 2007. p 12
Wagner C, Dulaunoy A, Wagener G, Iklody A (2016) Misp: The design and implementation of a collaborative threat intelligence sharing platform. In:

Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security. ACM. pp 49–56
Wheeler DA, Larsen GN (2003) Techniques for cyber attack attribution. Technical report. Institute for Defense Analyses, Alexandria VA
Woodier JR, Zingerle A (2019) The internet and cybersecurity: taking the virtual fight to cybercrime and cyberwarfare. In: Handbook of Terrorism and Counter Terrorism Post 9/11. Edward Elgar Publishing, Cheltenham
Yadav T, Rao AM (2015) Technical aspects of cyber kill chain. In: International Symposium on Security in Computing and Communication. Springer. pp 438–452
Zimmermann C (2014) Ten strategies of a world-class cybersecurity operations center. MITRE corporate communications and public affairs, Bedford. Appendices

Publisher's Note