Abstract: In times, where hacking back is increasingly considered as a legitimate reaction to cyber attacks against nation states, misattribution may undermine a state’s credibility and lead to political differences. Cyber attribution at this level must deliver reliable results. In recent years, threat intelligence services have often raised concerns regarding the reliability of attribution, and repeatedly pointed out the possibility of false flag operations. The intention of false flag campaigns is not necessarily to trick intelligence services but also to form public opinion. Unfortunately, there is a lack of a reliable approach that deals with the interdisciplinary challenges of cyber attribution. Additionally, there is a lack of concepts designed to deal with possible false flag operations on the technical side (e.g. manipulating digital evidences) and socio-political side (e.g. distributing fake news). Therefore, we propose a novel concept, the Cyber Attribution Model (CAM) to address these aspects. The model is divided into two closely interacting parts: Cyber Attack Investigation and Cyber Threat Actor Profiling. The scope of the CAM is mainly on professional and organized cyber attacks, such as espionage or APT campaigns, and designed for application in national cyber security centres. This paper presents further a literature research and the attribution model, (1) which is adjusted to today’s challenges resulting from the information war, such as false flag operations, and (2) which supports security experts – from technical analysts to intelligence services – to master the attribution process on all levels. Finally, we demonstrate the application of the Cyber Attribution Model in context of a real-world scenario.

Keywords: cyber attribution, profiling, cyber investigation

1. Introduction

The legal perspective of cyber counter attacks, such as hack back (Holzer & Lerums, 2016), is one of the most discussed topics today (Ponemon Institute, 2015). Misattribution of cyber attacks at the national level may undermine a state’s credibility and lead to political differences. Threat intelligence services have often raised concerns regarding the reliability of attribution in recent years, and repeatedly pointed out the possibility of false flag operations (Kaspersky, 2016). However, the attackers have an upper hand to reach their targets, while staying anonymous (Goman, 2018) or acting under a false flag, the processes of attributing threats to actors must deliver reliable results.

Nation states have always used information operations to enhance their goals, as conflicts have never been limited to the kinetic warfare (Liang & Xiangsui, 1999). The widespread weaponization of digital vectors has become commonplace also in the political realm. For instance, nation states and non-state organizations use technological means to distribute fake news, propaganda and other directed-content. Therefore, the attribution model covers the analysis of so called Influence Cyber Operations. According to the NATO definition, these are activities undertaken in and through cyberspace and qualify as cyber attacks with the intention of influencing attitudes, behaviours, or decisions of target audiences. Target audience can be everyone depending on purpose, from forming the public opinion to making security analysts believe that another threat actor is responsible for an attack. The social and psychological aspects of cyber attacks cannot be neglected (CCIOS, 2018) – quite the opposite is the case as they become increasingly significant in cyber operations. Recent incidents, such as the TVSonde incident, the Sony Hack (Sullivan, 2015), the Winter Olympic Hacking (Dion-Schwarz et al, 2018), show also the need for an analysis approach of possible false flag operations.

The CAM model distinguishes two types of false flags, one applied in technical context and one in socio-political context (see Figure 1). Historically false flag operations were usually conducted at sea, when one ship used the flag of another ship before attacking. Therefore, it was called a ‘false flag’ attack. According to the NATO terminology, a false-flag is a diversionary or propaganda tactic of deceiving an adversary into thinking that an operation was carried out by another party (NATO CCDCOE). There is a wide range of misdirecting actions. For instance, source IP addresses can be changed by using chains of proxy servers or the TOR network, as well as language settings, agent strings, and false hints in malware code can be placed. On the social side, threat actors could impersonate other actors and generate fake posts and news under wrong identities. It is extremely complicated to get all the false flags consistently right. Therefore, 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 might in fact be odd. The presented CAM focuses mainly on targeted and sophisticated cyber attacks and covers additional social aspects and possible false flag operation for reliable
2. Attribution Guide: Background and State of the Art

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. Our Cyber Attribution Model (CAM) separates the two corner stones: cyber attack analysis and threat actor profiling and brings them together in the attribution phase. The following section serves as a guide in the wide literature and aims at collecting common denominators of this interdisciplinary topic.

One of the most known models is the Q-Model by Rid & 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, Hutchinson & Margulies, 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).

Regarding threat actor attribution, the Hacker Profiling Projects has the biggest volume (Chiesa, Ducci & Ciappi, 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 Data’(2016PPAEPT, 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, Pendergast and Betz, 2013).

There are some common challenges in the presented models. The complexity and interdisciplinarity (technical-social-political dimensions) of cyber attribution are the main sources of many challenges. Attribution requires more levels of analysis and the cooperation between different fields of work (security experts, law enforcement agencies). There is no holistic model that unites all the cornerstones of this complex topic, such as digital forensic, hacker profiling, legal issues, and there is no model that focuses on the credibility of the digital evidences. Therefore, the developed Cyber Attribution Model tries to address all these challenges. Hence, for law enforcement agencies already fighting with cyber crime, cyber investigation serves as a basis for the CAM. Cyber investigation is developed based on new requirements resulting from the dependency from technology in almost all areas of the society. Law enforcement agencies need to react quickly on such changes. It is an extended and tailored solution to the digital aspect of crimes today.

3. Cyber Attribution Model

The Cyber Attribution Model (CAM) is based on the requirements derived from literature research, and on requirements of national authorities for cyber security (NIS authority or Cyber Security Center) in accordance to the EU NIS Directive. Typical tasks of a Cyber Security Center are to monitor the cyber threat landscape, as well as collect and process different types of threat-related information. The center has in best case cyber security experts with a solid knowledge on digital forensics, network and application security and penetration testing; law enforcement officers and intelligence officers. The developed CAM unites all relevant components from the technical analysis to threat actor profiling.

3.1 Overview
The biggest challenge is to create a universally applicable model. The findings in a cyber investigation arrive usually in a unique, unpredictable order. Hence, strictly defined and sequential flowcharts are not fitting the needs for cyber attribution processes. The application of the model is flexible, and the process flow can differ depending on the currently available information, findings and evidences. The Cyber Attribution Model (CAM) 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 Figure 1). Each part consists of Technical and Socio-political Contextual Indicators and the components of our reconsidered Diamond Model.

The primary aim of the Cyber Attack Investigation is to answer the questions, Who is the victim and Why, What has happened and How. Answering these questions is guided by the components: Victimology, Infrastructure, Capabilities 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 What 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 parallel in order to find a match between the applied TTPs and possible perpetrator profiles.

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 & 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 malwares, timestamps, strings, debug path, metadata, infrastructure and backend connection, tools, coding, language settings and pattern-of-life (Bartholomew & Guerrero-Saade, 2016). The difficulty to manipulate or fake technical indicators greatly varies, depending on the infrastructure of the victim and perpetrator. Please note that a detailed description of manipulating technical indicators would go beyond the scope of this paper.

![Figure 1: Cyber Attribution Model](image_url)
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 & 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 cyber space as a tool to shape the social and political mindset (Cohen & 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. The following section gives more details about the Cyber Attack Investigation and the Cyber Threat Actor Profiling.

3.2 Cyber Attack Investigation and Cyber Threat Actor Profiling

The Who: The starting point of the Cyber Attack Investigation is the victim. According to the retrospective view in Part I, the experts already know who the victim is, and they try to reconstruct the events. This happens by examining the victim itself. At this point victimology comes into effect. Victimology in the cyber space is found in the literature primarily 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 & 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 (Part II). 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 the fundamental practices of democracy. Ukraine, France and the United States were affected by attacks during their elections, for instance.

The What: The analysis of the Technical Contextual Indicators and the victim’s infrastructure help to reconstruct the operations. The Why: The analysis of the Socio-political Contextual Indicators help to understand possible motives for the attack. The How: After that, the required capabilities can be delivered. What are 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 to the incidents. Further results are TTPs, potential false flags on both side, and theories about possible opportunities and motivation.

Deriving TTPs answers mainly the What and How, helps to find the potential threat actors and to prevent similar attacks. The term TTP is often used in conjunction with Threat Intelligence, in the Structured Threat Information eXpression (Barnum, 2012). TTPs refer to Tactics, Techniques and Procedures and represent the behaviour or the modus operandi of the perpetrators. The term has its origin in the traditional military sphere and is used to characterize what an adversary does and how they do it in increasing levels of detail. There is a blurred line between the components of the TTPs. The CAM uses a definition, which we derived from the literature. 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. The activities of the APT threat actors are closely linked to the applied malwares and tools. As a consequence, there is a wild mix of naming conventions for APTs. Some use the name of the suspected threat actors, others use the name of applied malware or tools, which are sometimes also inconsistent due to their quick evolution. To sum up, TTPs are used to describes an approach of threat actors, and finally also well suited to profiling threat actors.

Findings from the investigation will be compared to potential threat actor profiles. Profiling is a delicate balance 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). The Cyber Threat Actor Profiling aims to create, update and manage threat actor profiles. It is the complementary part to victimology and helps to better understand what type of threat actor the perpetrator could be. Part I results in minimum required capabilities and observed TTPs. The analysts compare these results to known threat actor profiles. The What: The analysts are looking 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. The Why: The analysis of the motives of the threat actors, such as political motives for hacktivist groups or ideological motives for certain hacker groups. The How: The analysis, that the potential threat actor have the required components at all, such as access to confidential documents or to a certain code. 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 next section illustrates the application of CAM more visibly.

4. Application of Cyber Attribution Model

The selected scenario, the well-documented TV5Monde hack, serves as a basis for a short presentation of the application of the CAM (see Figure 2). The victim organization and the French intelligence services have shared their experiences about the incident. The application of CAM starts with Cyber Attack Investigation, especially with victimology. The victim is the TV5Monde, a French television network claiming to be one of the top three most available global television networks internationally (Oaktion, 2016). TV5Monde was a victim of a cyber attack, which caused service disruptions for hours in April 2015.

The circumstances need to 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 (EuObserver, 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-ISIS messages appeared on television’s Twitter and Facebook accounts. One of the messages posted 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.

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 is happened. 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 configured with a default username and password. Since the server was not connected to the internal network, the attackers continued the reconnaissance. They returned, using a compromised third-party account that allowed them to connect to the TV5Monde VPN on 6th February. After that, attackers began scanning internal machines connected to an infected endpoint and identified two internal Windows systems, that were used to manage cameras. Afterwards, the attacker used one of these compromised systems to create a new administrator-level user in the Active Director called “LocalAdministrator” on 11th February 2015 (Oaktion, 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 as much related information as possible, including information from the IT department’s internal wiki, which provided details on how logins and passwords were handled (Schwartz, 2017). After that, the attackers compromised another administrator machine with a Remote Access Control software, that was used for the sabotage. At 19:57 on 8th April, the attackers performed their first damaging operation by re-configuring 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 which 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. The investigations definitely showed the application of Sednit (aka Sofacy) malware, 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. The challenge is that there is no direct connection between the Pawn Storm campaign and the TV5Monde attack (TrendMicro, 2016).
These findings shape possible hacker profiles and so lead to the Cyber Threat Actor Profiling. To sum up the required capabilities, 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. The Cyber Threat Actor Profiling part supports to analyse, which potential threat actors can 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. In case, there are no fitting puzzle tiles, the analysts consider the possibility of false flag actions. At the time when the attack took place, the CyberCaliphate group initially claimed responsibility for the incident. The analysts however, had to examine, that the presumable perpetrator could have the required tools and infrastructure, the capabilities and the motivation to execute the attack. Despite the CyberCaliphate’s confession, the investigations revealed links to the Russian hacking group APT 28 (CFR, 2018). Finally, in case of the TV5Monde hack, there are three possible theories available with two potential threat actor profiles. A 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. The first theory is that TV5Monde was the victim of two entirely unrelated incidents, a Pawn Storm infestation and a separate
hacktivist compromise. The second theory is that the Pawn Storm group gave information, which was relevant for the attack, to a third party, 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. However, it is also possible that this attack was the work of undisciplined Pawn Storm actors. Though 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). 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 idea is that the Russian intelligence agencies go to cyber war against the West under an ISIS cloak (Observer, 2016). In that case, the attribution could piece the infrastructure, capabilities and motivation and the matching threat actor together. But since in this particular scenario, there were no clear technical evidences left, the analysts could have continued the attribution process using CAM. When there are no further digital traces, it is possible to continue with intelligence operations in order to underpin one theory with other evidences. As long, as there is no higher degree of certainty, the analysts cannot reliably refer to a potential threat actor and recommend possible (counter)measures. The Cyber Attribution Model aims to identify potential threat actors based on the findings in the cyber investigation considering possible false flag operations.

5. Summary and Future Work
This paper can offer only a brief insight into the Cyber Attribution Model (CAM) and into its application based on a real-world scenario. Its aim is to steer a reliable cyber attribution process which is adjusted to today’s challenges resulting from the information war, such as false flag operations. The main contribution of this paper is the new attribution model which supports the security experts to ask the full range of relevant questions, to aid their critical thinking and to put the investigation into a context. The future work contains the detailed description of the whole Cyber Attribution Model, the required cooperation between different fields of work (from technical analysts through law enforcement agencies to intelligence services), and a deep analysis of possible false flag operations from technical and socio-political aspects.

Acknowledgements
This study was partly funded by the Austrian FFG research program KIRAS in course of the project ACCSA.

6. References


