Public Attribution of Cyber Intrusions

JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2020, 1–12
doi: 10.1093/cybsec/tyaa012
Research paper

Research paper

Public attribution of cyber intrusions

Florian J Egloff [1,2]*

[1]Center for Security Studies (CSS), ETH Zürich, Haldeneggsteig 4, IFW, 8092 Zürich, Switzerland and [2]Centre for Technology & Global Affairs, Department of Politics and International Relations, University of Oxford, Manor Road, Oxford OX1 3UQ, UK

*Correspondence address. E-mail: florianegloff@ethz.ch
Twitter: https://twitter.com/egflo (@egflo)

Received 15 February 2019; revised 29 April 2020; accepted 19 May 2020

Abstract

Attribution is central to the debate on how to respond to cyber intrusions. The policy challenge is increasingly moving from identifying who is behind a cyber intrusion to finding the adequate policy response, including whether to publicly attribute. The article examines the use of public attribution as a political strategy for attaining specific political effects beyond the dyadic attacker–victim relationship, including shaping the operational and normative environment of cyber operations, with the potential to exert an independent deterrent effect. My analysis unfolds in three parts. The first part introduces two core concepts—sense-making and meaning-making—to capture different parts of the attribution process. I then introduce a theoretical understanding of public attribution drawing on the literature on revealing covert activity and argue that public attribution can serve the function of defining a particular interaction order, i.e. shape the rules of the 'game'. In part two and three I discuss two empirical examples of both concepts. I bring to the fore three observations: First, some states have shifted their policy responses from dealing with individual cyber intrusions to responding in a broader political framework of relations with a specific adversary leading to campaign-like responses. Second, the political decision whether to attribute publicly is not only a signal to the adversary, but also aims at shaping the future political and normative operational environment. Third, such norm shaping has the potential to exert an independent—though limited—deterrent effect, particularly on potential adversaries. The analysis demonstrates the importance of the meaning-making process to understanding the politics of attribution and the rewards of theoretically integrating it into the politics of secrecy and exposure of covert activities of states.

Key words: cyber security; attribution; public attribution; signals intelligence; national security policy; intelligence policy

Introduction

Attribution has been central to the debate on how to respond to cyber intrusions. Identifying who did it, and the uncertainty thereof, is one of the most discussed questions in the literature on cyber operations [1–11]. While these uncertainties are paralleled in many other realms of conflict, given the technologically driven and imperceptible nature of many cyber intrusions, not knowing who did it has been one key aspect in analysing the impact offensive cyber capabilities may have.

Though there had been research on attribution before (see e.g. [1–6]), in 2014, in a seminal contribution to the attribution literature, Thomas Rid and Ben Buchanan argued that 'attribution is what states make of it' [7]. They gave an encompassing overview of the attribution process and explained it from the bottom-up: from the technical forensics, over the all-source intelligence process, to its strategic probing and communicating the outcome. By outlining the attribution process as a whole, they charted a map of attribution activities and opened many doors for follow-up research. This

article picks up where Rid and Buchanan left off, by walking through one of those doors: the strategic element of attribution, particularly, the deliberate public attribution by national security policymakers.

This area is ripe for research, as the attribution landscape has evolved since Rid and Buchanan's article in 2014. Since then, strategically relevant state-sponsored cyber campaigns are frequently attributed to specific state actors.[1] Despite this, there is hardly any literature theoretically integrating this phenomenon in cybersecurity. This article addresses this gap by making a number of contributions.

To start with, I conceptualize what public attribution is. In order to do so, the article makes a conceptual clarification to disentangle two separate, but linked, processes in attribution: 'sense-making' and 'meaning-making'. Shortly defined, the 'sense-making process' in attribution refers to the knowledge-generation process that establishes what happened, whereas the 'meaning-making process' refers to deliberate actions that influence how others interpret a particular cyber intrusion. Using those two concepts, the article makes three further novel contributions.

First, I draw out a theoretical motivation of why (or why not) to publicly attribute. To do so, the article theoretically embeds public attribution into the literature explaining the revelation of covert activity. In line with that literature, it argues that one function of public attribution is to establish and uphold a specific kind of interaction order, often referred to as establishing the rules of the 'game'.

Second, empirically, the article argues that for some states, the 'attribution problem' has shifted into a policy problem of what to do about the cyber intrusion, including whether to attribute publicly. This is witnessed in the policy attention not being hung up on the technical attribution question 'who did it?', but rather on the policy question of 'what do you do when you think you know who did it?' That is not to say that the sense-making aspect of attribution, i.e. establishing—usually in private—what happened, is easy and always possible with a high-degree of confidence. The empirics presented show, however, in the cases investigated, sense-making was not the main challenge for policymakers, but rather the process of finding an adequate policy response. In some of the cases observed, all with developed national security policy capabilities, attribution is normalized and folded into the regular national security policy processes demonstrating that at the strategic level, the attribution process in cyberspace is not unique.

Third, building on the conceptualization of 'public' attribution as a specific kind of 'meaning-making' process,[2] the discussion of public attribution extends the current literature in three ways: First, the purview of the analysis is expanded beyond the dyadic relationship between the intruder and the victim, which brings multiple aims into scope for which public attribution may be undertaken. Second, in some of the cases investigated, public attribution is undertaken to shape the operational and normative environment of cyber operations, thereby establishing and stabilizing a particular kind of interaction order. Third, through so doing, the article argues against the strict necessity of punishing action for public attribution practices to have the potential of generating independent—though limited—deterrent effects, particularly on potential adversaries beyond the attacker in question [7, 12].

Methodologically, the article uses empirical examples rather than case studies, as its contribution lies not in theory testing, but in conceptual advancement. In order to do research on a specific issue, one first has to identify what kind of activity the issue of interest actually is [13, 14]. The article answers that question for the process of public attribution.

The analysis unfolds in three parts: The first part discusses the relevant cybersecurity literature on public attribution and introduces the two concepts capturing different aspects of attribution processes: sense-making and meaning-making. The concepts are theoretically embedded, drawing on the literature dealing with the revelations of covert activity. Particularly, I adapt the framework of tacit collusion to the revelation of cyber intrusions and explore its utility in better understanding public attribution. Note, we cannot yet rigorously adjudicate, whether and to what extent tacit collusion is operative: access to private data is too poor, the phenomenon evolves rapidly and in learning cycles, and a diverse set of political dynamics are driving it.[3] Nevertheless, this article contributes one possibility of how the phenomenon of public attribution can be integrated into existing theory, whilst recognizing that empirically, we currently can only keep abreast of necessary (but not sufficient) explanations for the phenomenon at hand. We can expect that the accumulation of knowledge slowly will lead towards a better grasp of it. The conceptual contributions and theoretical integration of the phenomenon in this article are a start to this endeavour.

The second part argues that at the strategic level, the process governing attribution of cyber intrusions is not unique. It observes that at least for some states, relatively rapid sense-making capabilities are at play, leading to questions on what to do when you have a medium- to high-level of confidence in your attribution judgements. To illustrate this empirically, the article traces these rapid sense-making capabilities in two empirical examples: the 2015 intrusions into the German Bundestag and the 2015–16 intrusions into the DNC.

In the third part, the article considers deliberate public attributions by national security policymakers as a political strategy, and observes their potential longer-term effects, particularly with regard to shaping the operational environment for future adversaries. It highlights this strategic approach to public attribution by examining two public attribution campaigns (WannaCry and NotPetya) and discusses their function in establishing the rules of the 'game'.

The conclusion summarizes the arguments, assesses their durability and limitations, identifies areas of further research and offers an outlook on the anticipated future developments of public attribution.

Part I: A theoretical understanding of public attribution

This part first discusses the literature on public attribution in the realm of cybersecurity. I then introduce 'sense-making' and 'meaning-making' to disentangle the attribution process. This expands on one of the elements discussed by Rid and Buchanan (communicating attribution) by making a conceptual clarification offering a better

1 Examples are the hack of the German Bundestag (2015–16), the intrusions into the United States Democratic National Committee (DNC) (2015–16), the WannaCry ransomware (2017), partially bringing to a halt the UK National Health Service, as well as NotPetya (2017–18), which the White House claims to be the most destructive and costly cyber-attack in history. (The first year refers to the start of the campaign, the last to the date of the public attribution.)

2 Particularly, deliberate public attribution by national security policymakers is identified as a political strategy for attaining specific political effects.

3 This is why I call my theoretical contribution 'a theoretical understanding of public attribution'. I thank one of the anonymous reviewers for clarifying this point.

understanding of one element of attribution processes: public attribution. The final section theoretically embeds public attribution drawing on the literature explaining the revelation of covert activity.

Public attribution in the cybersecurity literature

The literature exploring the possibility of public attribution in cybersecurity deals with it as an incident-based decision. Rid and Buchanan point out three benefits of publishing attribution judgements (improved credibility, attribution, and defences), highlight the risks of losing insights into adversary operations, and demand consequences if publicly attributing, lest the adversaries be bolstered in their actions [7]. Clement Guitton casts public attribution as a 'game to convince an audience', which depends on trust and authority of the entity making the claims [15]. Lin extends this line of argument by discussing different standards of evidence in the international and domestic legal and intelligence settings. However, he concludes that there are no such evidentiary norms effective at the international policy level [16]. All authors cast the public attribution part of the attribution process as an incident-based decision. In contrast, this article argues that the public attribution campaigns, witnessed in 2017–18, are an outcome of an integrated national security policy response. They are not 'just' reactive to one particular campaign, but an outflow of a broader national security policy. Consequently, whilst the attribution judgement offered may pertain to a single 'incident', the decision-making at the policy level is much broader than single intrusion-based considerations.

Timo Steffens discusses public attribution as an element of political strategy in more depth. He highlights the information gain/loss ratio of public attribution. This refers to the concept that disclosure of information can lead to intelligence/technical gains and losses, e.g., through gaining insights by seeing the adversary adapt to the disclosure, or losing collection on the adversary, because the adversary shifts outside of one's visibility. He also demands that each public attribution judgement be tailored to achieve a specific goal, be audience specific, and be communicated with a certain quality [12]. Protection, justice, diplomacy and politics, as well as—for companies—reputation are such high-level goals that can motivate an entity to attribute publicly.

Of interest here are Steffens' arguments regarding diplomacy and politics. He argues that for state agencies, public attribution statements are carefully weighed decisions, mostly intended to send 'symbolic or political signals' [12]. Since signals can be sent on diplomatic bilateral channels, public attribution should be understood as an escalatory step [12]. Steffens then points to both companies and other states as the relevant audiences, whose follow-up actions can increase diplomatic pressure. Thereby, it is important to identify whether the government just wants to send a signal to the adversary over a public channel. If so, the details of the attribution judgment need not convince a public audience, but rather the well-informed party that committed the intrusion. Just like Rid and Buchanan, Steffens argues for consequences to result from public attribution, as, for example, represented in economic sanctions or sanctions on personnel (e.g. *persona non grata* orders). He sees in them a deterrence practice, demonstrating the capability to attribute and the resolve to undertake sanctions against actual and potential adversaries [12. Also 7].

As reflected in Steffens' situating of the public attribution decision space, whether to attribute cyber campaigns publicly has become an element of national security strategy decision making. This article expands the analysis of effects of public attribution beyond the adversary in the specific case: public attribution is not only being aimed at the adversary, but at a broader audience. By doing so, it puts into question the conceptual necessity for independent punishing actions, in order to exert deterrent effects, as argued by both Steffens and Rid and Buchanan. Rather, and as will be argued below, public attribution itself can exert some, if limited, deterrent effect on a specific sub-group of potential adversaries. This reading integrates public attribution in a theoretical framework presented below, with a very particular function: public attribution can be used to signal one's interpretation of the rules of the 'game'.

Sense-making and meaning-making processes

The process of attribution can be split into two conceptual processes. The 'sense-making process' refers to the knowledge-generation process that establishes what happened. In this phase, the party driving the attribution process (e.g. a victim government) is trying to make sense of what happened. This process includes the communications amongst the entities who are involved in making an analytic attribution judgment. The majority of Rid and Buchanan's analysis focuses on detailing this process conceptually and empirically, as an integration of the tactical data analysis of incident response with the operational all-source intelligence process, and the concurrent strategic probing of its outcomes [7]. For larger cyber campaigns, the sense-making process includes giving an account of the organizational processes that led to the operational decision making by the adversaries. Moreover, for campaigns with a strategic political aim, an assessment of the sponsor's aims, their context, and stability over time is able to situate the behaviour in cyberspace in a larger political context of bilateral and multilateral relationships. Attribution judgements regarding responsibility, whether in the cyber domain or not, are intelligence judgements, which always come with uncertainties attached [e.g. 17].[4] Sense-making, thereby, is an internally focused process, where an actor is trying to approximate the baseline truth (for how information warfare aims to influence sense-making, see [18]).

This is opposed to the 'meaning-making process', which involves the 'production of facts, images, and spectacles aimed at influencing socio-political uncertainty and conflict generated by crises' ([19]; p. 148 [20]). The meaning-making process is aimed at communicating an attribution judgement to others, in order to change the uncertainty structures associated with the particular intrusion and to exert political effects. Both processes are often overlapping: the sense-making process of attribution is a continuous, usually secret, process, often driven by intelligence agencies. The meaning-making process is a national security policy process, involving a broader set of policy participants from across the government.[5] It is an externally focused process, in the sense, that the actor decides on what meaning-making activities to engage in, using the intelligence judgement arrived at in the sense-making process.[6] Public attribution is a

4 Importantly, they do not claim to be scientific. Rather, they are focused on enabling a particular customer to make decisions. Note: this is a different language use than international law establishes for its technical term 'attribution'.

5 I focus on deliberate public attribution by a government. There could, of course, also be accidental meaning-making activities.

6 Can a sense-making process unintentionally acquire meaning-making quality? Yes, it can. For example, when a fact about the current sense-making process leaks to the media. If this is unintentional, it would not qualify as a deliberate meaning-making process. I use unintentional, as if it is done intentionally, the two conceptual processes can again be distinguished.

special kind of meaning-making process, namely one, in which the means of engaging in the meaning-making are publicly visible. This is opposed to the private meaning-making activities, which a government may engage in, e.g., by using diplomatic démarches, victim notifications, or informing developers of vulnerable software. One would regularly expect governments to coordinate their responses in private, e.g. by informing victims, as well as allies, before publicly accusing another state.

Colluding and revealing covert activity

We can further unravel sense- and meaning-making from a theoretical perspective by drawing on the literature explaining the communication about other forms of state originated covert activity. Particularly pertinent in this regard is Austin Carson's work on tacit collusion [21, 22], which I will draw on and adapt to cyber intrusions. Tacit collusion is a form of backstage theatre, in which a different interaction order is operative than in publicly visible domain. An interaction order is a basic framework of rules and roles that defines the interaction space as a particular kind of game (p. 111 [20]). Tacit collusion thereby refers to state silence about knowledge of a covert activity undertaken by an adversary. This is partially operative in cyber insecurity. One of the main explanations why states tacitly collude was theorized to be the risk of undesirable escalation i.e. that if one revealed the activity, the state would enter a different interaction order and, consequently, have to commit major resources to responding to the activity.[7] To the degree that strategic interaction in cyberspace has been found to be less useful for coercion and less escalatory in nature, we could expect publicizing a covert cyber activity to also carry less risk of undesirable escalation [23, 24]. The escalation risk is further alleviated by the presence of security companies that may disclose an activity before a state does (p. 296 [21]).[8] The lack of overt escalation occurring through public attribution so far seems to confirm this judgement. Thus, this lower escalation risk could partially explain the presence of public attributions of cyber intrusions (i.e. why the incentive to uphold tacit collusion is lower than in its originally theorized context, that of covert military intervention).

So why do states publicly attribute? Public attribution serves to shape the operational space and aims, in part, to shape the interaction order, i.e. the rules and roles that define the interaction space as a particular kind of 'game'. This claim requires a longer explanation: Overall, offensive cyber capabilities are mostly rooted in signals intelligence. However, they have, due to the scale and depth of collection possibilities as well as the possible effects across targets, due to the commercial of the shelf technology being shared across different types of actors, transformed into a separate 'game', or more formally, an interaction order (for a strategic analysis arguing for a separate interaction order for cyber, see [25]; for a historical analysis of the transformation of the espionage 'game' through the addition of cyber, see [26–28]; for cybersecurity primarily being an intelligence problem, see [29, 30], also [31]; for cyber espionage norms specifically, see [32–34]). As in any game where no referee exists, the players make the rules through playing the game (for the practice theoretical approach to this see [35]).

Disagreements about the rules may persist for an extended period of time. Thus, this element leads to the expectation of public attribution serving not just to communicate the rules and roles to the adversary, but to also signal to the potential adversaries and allies one's interpretation of the rules of the 'game' (similar, see [36]). The public channel thereby serves the function of maximum reach: if you want to induce alignment not just of a direct counterpart, but the broader operational community in cyber operations worldwide, it is much more effective to do so publicly.

Thereby, theoretically derived collusion is reserved to those activities that either all parties agree that they are legitimately performed and hence kept secret (certain forms of espionage, see e.g. the discussions in the OPM case below; see further [37, 38]), or, that through their publication one would endanger other rules and regimes [39] or undesirably escalate outside of the agreed space of competition [22, 40].

This is worth further discussing in turn. First, activities that are legitimately performed and hence kept secret by all parties. A state that wants to use public attribution to shape the rules of the game, will less likely politically denounce other states for performing activity it itself considers legitimate (and, perhaps, it engages in itself). For example, the existing toleration of espionage leads to the expectation of tacit collusion around espionage-based intrusions. This does not mean that states never publicly attribute a cyber incident to another state for an activity they themselves consider legitimate state activity. However, an information gathering activity is much less likely to be publicly attributed to another state than an activity that would be considered illegitimate by the attributing party. This could be one of the reasons why states have more frequently attributed activity undertaken by the threat actor aliased APT28 than activity by the one aliased Turla, i.e. Turla regularly having a classic espionage profile vs. APT28 using cyber espionage as a first stage for more active effects.

Second, endangering other rules and regimes by publicising activity: Because there currently is only a very thin normative system (e.g. UN GGE non-binding rules or OSCE confidence building measures), much less a formalized framework, governing activities in the cyber domain below the threshold of armed conflict, there exists little risk of endangering cyber rules and regimes (see Carnegie and Carson in [39]). The current frameworks outline broad voluntary rules, but there is no enforcement attached to it. However, where publication could endanger other (non-cyber) treaties or regimes, a state may be reluctant to publicly attribute an incident, for fear of public evidence of non-compliance leading to more non-compliance by other states. This could, e.g., be the case with states that are not publicly known to pursue economic espionage. A state wanting to build (or uphold) the regime against economic espionage may be incentivized to not publicize such evidence, for fear of fuelling non-compliance.

Finally, the fear of undesirably escalating outside of the agreed space of competition could apply to some of the activity that sometimes is labelled 'preparation of the battlefield'. If, e.g., an actor is caught testing their capabilities of interfering with another state's military command and control systems, publication of such activity could potentially open up undesired escalation risk. Thus, collusion is more likely in such cases, because the victim state fears publication could escalate into a different space of competition.

7 Carson theorizes this as escalating out of one interaction order (limited war) and entering another interaction order (regional or global war).

8 Carson acknowledges both elements in his book: the 'lack of escalation risk, or the exposure of a sponsor's identity by third-party non-state actors, should make exposure more attractive as a tool to diplomatically isolate and punish the sponsor' (p. 296). One might add that non-cyber covert activities also get disclosed by third parties, such as media outlets. However, in the case of cyber activities, there exists a whole industry focusing on analysing covert cyber activities.

Due to the difficulty in sense-making in cybersecurity, the gain/loss ratio weighs heavily when disclosing attribution judgements, leading to a pronounced disclosure dilemma [41]. This is because the adversary faces an uncertainty distribution around discovery. Thus, the defending actor cannot by default expect that the adversary assumes visibility on the defender's behalf (as e.g. assumed in [21], see p. 49). Only, in rare cases self-attribution may be of interest to the adversary [42]. Some operations are cloaked under implausible deniability, a concept that refers to apparent, but unacknowledged activity. Cormac and Aldrich explain that there always has been a spectrum of visibility and acknowledgement and specifically point to the multiple audiences as the cause of this spectrum: 'In reality, a spectrum of attribution and exposure exists since covert action has multiple audiences, both internal and external' (p. 479, [43]).[9] Some attackers may want exposure of the plots, as was the case with the KGB, which considered the exposure of forgeries an operational success, creating mirror imaged analyses, and create 'fear and uncertainty about the present' (pp. 492–93). However, the authors also warn of blowback when abundantly relying on covert action, particularly the feeding of conspiracy theories creates the danger of uncontrollability of effects and through that, the potential for escalating to conflict (p. 493). Thus, to the extent the adversary is interested to operate in the visible, but unacknowledged space, implausible deniability is also operative in the cyber incidents. Public attribution in this case is desired by the perpetrator.

The added complication of sense-making in the cyber domain increases the significance of the intelligence gain/loss ratio in the decision to publicize. States publicly attributing are facing a benefit and cost from revealing information. In other domains, this disclosure dilemma was bridged in disclosing to an international organization, rather than to share with the public [41]. No such international organization exists in cybersecurity, and in its absence, research would expect states to underreport 'violations of international laws and rules' [41]. Even if those rules, as argued above, are still unclear, states judging other states to have violated the rules of the 'game' are stuck with either disclosing hard won intelligence or disclosing just their conclusions, but face a domestic and international credibility problem, i.e. the problem of convincing an audience of conclusions without presenting conclusive evidence. This can be partially overcome by internationalizing the disclosure in attribution coalitions. This serves as a second-best option to solving the credibility problem through an international organization (on non-state elements of these decentralized attributions, see [44–48]).

To summarize: Tacit collusion is partially operative in cyber insecurity. Due to the difficulty in sense-making, the gain/loss ratio weighs heavily when disclosing attribution judgements. Theoretically derived tacit collusion is reserved to those activities that either all parties agree that they are legitimately performed or, that through their publication one would endanger other rules and regimes or escalate outside of the agreed space of competition. In particular, I argue, that because the (political) rules of the 'game' are still unclear in cyber operations, public attribution is used to shape the operational environment with the aim of establishing and sustaining such rules of behaviour, i.e. stabilize a particular interaction order. It can thus be read as an affirmation of a particular interaction order. This norm shaping activity can be recognized as an independent element of public attribution. It has the potential to exert an independent deterrent effect on a sub-group of adversaries, irrespective of other consequences being imposed. Thereby, the attribution coalitions serve the purpose of bridging the credibility problem at a domestic and international level.

Part II: The attribution problem understood as a sense-making process

Having introduced sense- and meaning-making and theoretically embedded it in the literature, this part empirically focuses on the sense-making aspect of attribution. As laid out above, this constitutes the aspect that traditionally is captured by the cyber specific literature as the 'attribution problem'. The sense-making process is discussed, with particular emphasis on its generic quality comparing it to the domain of chemical warfare. Two empirical examples, the Bundestag and DNC incidents, illustrate how sense-making in the cyber domain can take place relatively rapidly. The two cases are relevant, as they both represent strategic events for the respective polities: if the 'attribution problem' were as pernicious as the cyber literature portrayed it, we would expect it to also have a bearing on strategically relevant incidents. We will see that not sense-making was the main challenge in the cases investigated, but rather the policy challenge of what to do about it. Public attribution forms part of those policy response options.

In a second step, in part III, the strategic logic of public attribution as a meaning-making process are discussed in detail, with a focus on its relationship to tacit collusion. Two empirical examples, WannaCry and NotPetya attributions are used to illustrate three logics operative in public attribution: (i) shaping of the operational space, (ii) the developing and cultivating of norms and (iii) the limited deterrent value created through such actions. The two empirical examples are relevant, as they form the best illustration for how the logic of the argument to applies. Thereby, the discussion is brought back to the tacit collusion framework, in particular, to the function of public attribution in signalling the rules of the 'game'.

For the empirical illustration of the argument, it is important to consider the problem of non-observance, a problem not unique to the area of cybersecurity. Rather, it warrants the generic disclaimer often made in research on covert activities, namely, that we know that we are only observing partial data. We know this is the case from private intelligence reporting as well as public reporting on previous incidents that we did not know about at the time (see e.g. Ron Deibert on the notification gap in [44]). The important question for research on public attribution is whether the type of data used to illustrate the argument is representative for the phenomenon of public attribution overall. The argument that public attributions can be undertaken with the strategic aim of shaping the operational space, develop and cultivate norms, and generate limited deterrence value, thereby rests on the cases selected illustrating that this was one of the aims. Thereby, it is not argued that this is the only possible motivation for public attribution, ameliorating our problem of representativeness, a point that will be elaborated in the third part of the article. Overall, we would find less confidence in the argument if data was presented to the effect that this was not one of the considerations. Confidence in the argument is strengthened, when identifying cases, where public attribution is not undertaken, for reasons of undermining the desired

9 However, their focus of research is mostly on the initiator of the covert action, not on the defender. They argue that operating in the 'grey zone between secrecy and exposure brings significant benefits' (p. 493). These include the sub-escalatory demonstration of resolve, creating uncertainty and 'fear abroad and yielding electoral benefits at home' (p. 493).

operational space, normative frameworks, or deterrent logics. At this stage, it is too early to make broad generalizations across the entire public attribution phenomenon. The best one can do is to identify different categories of public attributions and identify the strategic logics at play.

Attribution problems are not unique to cyber security

The attribution debate in cyber security is quite peculiar: much of the literature fetishizes technical evidence, whether its absence or its presence, as if it alone could 'prove' who is responsible. This is particularly true in one sub-strand of the technical debate, where presenting technical evidence is a precondition for a statement to be believed at all. For example, a column in *Computer*, where the politics of public attribution are posited as 'faith based attribution' and strong scepticism of any statements made by the government, which are not backed up by public technical evidence, are voiced [50]. There is no discussion of the reputational loss a democratic government were to suffer if it purported claims that are later uncovered as lies. That is not to say technical evidence is not important, but it on its own does not explain, which social entity is responsible for an action. Other sub-strands of the technical debate agree with this. For example, Guerrero-Saade and Raiu argue that cross-domain attribution is 'beyond private sector fifth-domain capabilities' but within reach for intelligence agencies, who rely on their 'visibility *within and* ***beyond*** the fifth domain' [51].

The attribution problem in cyberspace is not unique. Technical evidence is not sufficient to assign responsibility in other domains of international conflict either. Other contexts are just as much fraught with the interplay of highly technical, material evidence, and inference processes that lead to the attachment of an action to a particular actor. In some contexts, those judgements are highly regularized. For example, in many jurisdictions, in the context of a court procedure, the evidentiary value of DNA typing is a relatively settled matter—this is despite criticism pointing to problems in the use of DNA typing [52]. In other contexts, those judgements are highly politicized, and no stable expectation around evidentiary standards exist [16].

The chemical poisoning of Sergei Skripal in 2018 is an instructive case from the domain of chemical warfare. After the chemical attack against Sergei Skripal, the UK formally asked for technical assistance from the Organisation for the Prevention of Chemical Weapons (OPCW). The OPCW relies on its associated laboratories, some of the best chemical laboratories in the world, to identify substances scientifically. They can establish what substance was used and potentially the amount found at the scene. The inspection team may also identify 'any impurities or other substances' which 'serve to identify the origin' of chemical weapons analysed [53]. However, responsibility claims are not established based on 'technical' evidence. Rather, they are the result of the sense-making part of attribution, an all-source knowledge generation process, in which some account of human decision-making and organizational structure explains the behaviour and technical evidence observed. Thereby, the technical evidence is the result of human behaviour, not the behaviour itself.[10] In the domain of chemical warfare the element of identifying the responsible parties has evolved, with the OPCW-UN Joint Investigative Mechanism (OPCW-UN JIM 2015-2017) being explicitly tasked to identify responsible parties. Subsequent to the Russian veto of the renewal of the OPCW-UN JIM and the Skripal poisoning, the UK successfully proposed to strengthen the OPCW Secretariat's role in attribution processes—a change that was adopted by a majority vote [54, 55]. Crucially, the determination of responsibility in the domain of chemical warfare is the result of an all-source knowledge generation process [56].

Observing sense-making empirically

In recent years, many states have prioritized and invested in both offensive and defensive cyber capabilities, including building expertise and capabilities in the area of attribution. Herb Lin locates one reason for the improvement in attribution capabilities that 'more people are paying attention' [16]. More organizations are willing to prioritize and invest in attribution capabilities, anticipating more malicious cyber activity in the future.

The empirical record enables us to observe how these sense-making processes take place. At the national security policy level, some states, e.g. USA and UK, have shifted from an incident-driven approach of attributing a single breach to pro-actively observing entire intrusion campaigns. Empirically, the attribution processes in states have evolved. In some cases, the question of who is behind an intrusion is being answered relatively swiftly (within weeks, not months, sometimes even swifter).

Importantly, if the sense-making process is at a stage of reaching a medium-to-high confidence judgement, the attribution process is not finished. Rather, as Rid and Buchanan pointed out, communication is part of the attribution process. As witnessed in several cases, the policy process between having robust knowledge of who is behind an intrusion campaign and the decision to take policy action based on such knowledge is what ultimately resulted in several months of inaction. This policy process, thereby, is not unique to cyber intrusions. Some states have made investments into the attribution policy process. Cyber intrusions have been integrated into the national security policymaking, with all the constituencies and equities attached. Thus, the response to the intrusion campaigns has become integrated into a larger framework of interaction with the adversary at hand.

Two empirical examples are instructive as indicators of these rapid sense-making processes: the Bundestaghack of 2015 and the intrusions into the US DNC in 2015/6.

In late April 2015, members of the Bundestag, the lower chamber of the German parliament, were targeted as part of a phishing campaign [57].[11] By 19 June 2015, an independent report by security researcher Claudio Guarnieri (sponsored by *Die Linke*) concluded that the technical artefacts and analysis 'suggest that the attack was perpetrated by a state-sponsored group known as Sofacy (or APT28)' [58].The report also detailed the technical evidence gathered, including configuration details of two pieces of malware. It made forensic links to previous APT28 campaigns. Despite the public report being available, the German government did not publicly comment on the attribution of the espionage operation. Secretly, however, the government independently arrived at the same conclusion. On the 2 July 2015, a representative of the federal office for information security (BSI) informed the parliamentary committee in charge of analysing the intrusions: 'the attack pattern and the exfiltration of data is consistent with, the from other sources

10 The use of autonomous agents in offensive operations may transform accountability claims.

11 Note: The email reused a title of an article published in the Daily Telegraph on the day before, the 29 April 2015. Corroboration of the spoofed United Nations origin of the e-mail is provided in Even, B. Propaganda Und Politische Einflussnahme Als Strategische Handlungsoption Ausländischer Nachrichtendienste. In: *Neue Gefahren für Informationssicherheit und Informationshoheit: 10. Sicherheitstagung BfV und ASW Bundesverband*, Berlin, 2016. pp. 24–32. Bundesamt für Verfassungsschutz, Berlin, Germany, 24–32.

known, APT28' [59].[12] Thus, since at least 2 July 2015, the BSI was confident enough in its assessment to share it with representatives outside of its own agency, demonstrating a German 'sense-making' capability.

Its 'meaning-making' capability, however, was still in development. It took the government another 6 months until the first visible policy response against the adversary (the start of a prosecution against an unknown party), and a full 10 months until the release of a public statement officially attributing the intrusion to the Russian Federation [60, 61]. During that time, the German Chancellor pushed the national security community (including the intelligence agencies, the foreign office, interior ministry, and the ministry of defence), to assess the background of Russia's confrontational actions (within and outside of cyberspace), indicating a will to integrate cyber intrusions in a broader deliberative national security process [57]. Only years later, in 2020, the German Chancellor went on the record calling the incident 'egregious', confirming that the federal prosecutor general is searching for a Russian suspect, and stating that the government reserves the right to implement further measures, also against Russia [62].

The case of the intrusions into the DNC before the US presidential election in 2016 is a second example for both the rapid 'sense-making' capabilities and the policy responses presenting the main strategic challenge. In April 2016, Crowdstrike was called to help with incident response for the DNC. They discovered at least two threat actors were active on the DNC networks, one had been there since at least summer 2015, the other since April 2016. On 14 June 2016, the DNC, together with the cybersecurity firm Crowdstrike, publicly attributed the intrusions into its networks to two separate Russian espionage groups, APT28 and APT 29 (or FancyBear and CozyBear in Crowdstrike's terminology). By that time, the White House was already tracking several Russian activities connected to the US campaigns.[13] By late July 2016, the cybersecurity community had identified parts of the broader Russian subversion campaign, including some of the influence elements [63–66]. The latest in early August 2016, the US President Barack Obama was briefed on the political intent of Russian interference and on the Russian leadership's direct involvement [67]. Despite the broad sourcing and unusual clarity of evidence, it took another month for members of the legislative to go public, and 2 months for the US government to release a meagre public statement by the executive. Only 5 months later, 3 months after the elections, the government released an intelligence community assessment attributing various interference activities to the Russian government [68–70].

Much of that delay in the US public response is explained by domestic political concerns of the executive being seen to unduly intervene in the election process. At the domestic level, the procedures were underdeveloped and unprepared for a coordinated response between the federal government and the state-based election officials (on the different roles of the state, see [71]). Inter-agency discussions and indecisiveness about what to do next significantly slowed down this process. For example, in August 2016, then FBI Director James Comey drafted an op-ed publicly attributing the Russian activities that was debated in the inter-agency process but was never published [72]. Even so, having had experience with the policy discussions surrounding public attribution, the US policy process was one of the more developed for the international dimension. The policy discussions reveal that the 'sense-making' did not delay the policy process. Rather, it was the question of the appropriate response that was on decision-makers' minds, namely, 'what to do when you think you know who did it?' The possible response spectrum is wide, ranging from doing nothing, to covert options, to very public confrontations, including becoming very active in the adversary's networks. Other authors have commented on why the more escalatory and symmetric response spectrum was capped off in this particular instance (i.e. why the USA was 'deterred' from more aggressive responses) [73–75]. Of note, at the end of July 2016 there were some indications of a possible US covert cyber response. The FSB issued a press statement of the government suffering a widespread intrusion ('a targeted virus spread, planned and made professionally') [76].[14]

Both cases demonstrated rapid sense-making processes at play. Public claims by the private sector thereby aided the reduction of uncertainty in public discourse about the background of the intrusions. The cases showed that sense-making was not the main challenge for the strategic response to the cyber intrusions. Rather, the focus was on finding the adequate national security policy response. One significant aspect of such a response is whether the government decides to attribute publicly, and thereby actively engage in meaning-making processes to shape the interpretation of the intrusions and achieve political effects with several stakeholders. Part III addresses the strategic considerations of that course of action using two empirical examples: WannaCry and NotPetya.

Part III: Public attribution as a political meaning-making process

This third part argues that beyond a situational response, deliberate public attribution by policymakers is always a political strategy.[15] It situates the decision whether to attribute publicly not as an incident-based decision only. Rather, recent empirical cases exhibit a more long-term, coordinated approach (WannaCry, NotPetya), linking cyber (in-)security to national security strategies more broadly. Such a longer-term approach focuses on shaping the operational environment of adversaries, aiming to establish rules of behaviour, and exerting a norm-shaping effect. This could have the potential for a deterring effect over the long-term.

Public attribution as an element to shape the rules of the 'game'

The subsequent argument can be broken down into three pieces: first, some states do not deal with cyber intrusions as incident-based decisions, but rather contextualize them in a broader framework of relations with the specific intruder, focusing on identifying the strategic intent of adversary campaigns of (cyber and non-cyber) activities, and identifying possible policies to respond to such an intent.

12 Literal translation by the author. The original German quotation is: 'Das Muster des Angriffes und der Ausleitung von Daten entspräche dem bereits von anderen Stellen bekannten APT 28'.

13 At working level, the FBI had the DNC intrusions on the radar since 2015, but there is no public evidence of it briefing the White House before 2016.

14 If the USA was behind it, then the head of NSA's TAO giving a rare interview at the Aspen Security Forum about the NSA 'hacking back' adversaries for attribution purposes, published a few hours after the FSB press release, could also be read as a signal of taking credit. See L. Ferran, The NSA Is Likely "Hacking Back" Russia's Cyber Squads, *ABC News*, 30 July 2016, https://perma.cc/27XT-3G4K.

15 An analysis of the 'private' (i.e. diplomatic) attribution claims is out of scope for this article.

Second, the political decision whether to attribute publicly is not only signalling to the adversary, but also shapes the future operational environment, particularly when the aim is to establish rules of behaviour, i.e. to establish and stabilize a particular interaction order. Third, over time, such a norm shaping effect has the potential to exert an independent deterrent effect, irrespective of other consequences being imposed.

Whilst individual incidents become the focus of public discourse, many (if not most) cyber intrusions never surface to the level of public visibility.[16] Consequently, when observing state responses with a public element, such as deliberate public attribution, it is likely that the state is responding not just to the individual incident, but also to broader campaign of activities. This judgement is supported by two empirical observations. First, cyber intrusions themselves are often organized in campaign-like structure.[17] Second, particularly state-led cyber intrusions do not take place in isolation, but are integrated in a broader strategic intent manifesting itself across other activities of an adversary. The level of integration is dependent on the exact nature of the activity, the size of the government, and the level of coordination between different entities in the government. Coordination is sometimes recognizable, e.g., by the concurrent activities of other ministries and cyber intrusion activities. Assessing the strategic intent is crucial in order to tailor an effective response, despite the difficulties posed in the cyber domain [77]. Of note, the decision-making on public attribution should not be seen in isolation either. Just as the adversaries organize their activities to fulfil a strategic intent, defenders also use different parts of their toolkit to attain their strategic aims.

The political decisions of whether to attribute publicly thereby face many considerations, many of them raised by Steffens [12]. The one highlighted in this article is the possibility to not only send a signal to the adversary, but focusing on longer-term effects of shaping the environment, with the aim of establishing rules of behaviour. This is witnessed in two diplomatic initiatives launched by the Five-Eyes signals intelligence collaboration.[18] Particularly, in response to the WannaCry ransomware in 2017, the UK, whose national health service suffered severe outages, led a diplomatic campaign to attribute the ransomware publicly to North Korea, resulting in official public attributions by all the Five-Eyes countries, Denmark and Japan [78]. Of note, the government-initiated campaign came with a delay of 7 months between the intrusions being attributed by the private sector and the government going public [79].[19]

Eight months after, NotPetya,[20] another severe destructive cyber campaign aimed primarily at Ukraine (but spreading quickly worldwide), the same group of states (without Japan, but with Ukraine) again decided to attribute the activities publicly, this time to the Russian military.[21] The UK's attorney general, Jeremy Wright, outlined the reasoning behind this in a speech at Chatham House:

> If more states become involved in the work of attribution then we can be more certain of the assessment. We will continue to work closely with allies to deter, mitigate and attribute malicious cyber activity. It is important that our adversaries know their actions will be held up for scrutiny as an additional incentive to become more responsible members of the international community [80].

This joint approach of the Five Eyes was corroborated by the US state department and the US national cyber strategy, which highlighted the importance of partners in buttressing each other's attribution claims and responses [81–83]. Thus, the Five-Eyes governments have decided that they use public attribution as a way of shaping the environment. By building an international coalition, attributing blame publicly to an actor for specific actions that are deemed undesirable to the international community, the coalition of states is changing the operational environment, and in Carson's terms, is delineating the rules of the 'game'. The intent, as outlined by the UK government official, is to establish boundaries for responsible behaviour, respectively clearly labelling the behaviour deemed 'irresponsible'. Indeed, the United Kingdom hopes to attain benefits with pursuing public attribution more broadly, including making cyberspace 'more transparent as counter-normative and destructive behaviour (i.e. Wannacry and NotPetya) are attributed' leading to 'greater stability in cyberspace as clear lines of unacceptable behaviour are drawn', increasing the legitimacy of attribution by undertaking attribution with allies, and using attribution as a first step 'to enable wider response options to impose costs on the responsible actors' [84]. Thus, the British government identifies its activities in public attribution as boundary drawing behaviour that gains more legitimacy as it is undertaken in coalitions of states. This is in congruence with the literature on international norms: in order for norms to be effective, norm violations have to be acknowledged [85, 86]. The norm shaping activity seems to be targeted at encouraging state actors to not indiscriminately deliver effects. Though an in-depth international legal analysis is out of scope, one can note, that by acting in coalition states prepare the ground for establishing state practice, one of the sources for international customary law (see further [87]). However, in order for this to develop into standing state practice, it would be important for states to refer to the specific rules of international law breached. Without it, the statements may be important for norm setting, but do not in themselves have any international legal quality and thus remain political condemnations [88].

Public attribution in the form of an attribution coalition can also be read as a form of overcoming the credibility problem. Just as Carnegie and Carson have laid out using an international organization in the non-proliferation space, in its absence, in the equally sensitive area of cyber intrusions states may be trying to build credibility through leveraging trust in various governments versus

16 As discussed in part II, the issue of non-observance, and the associated missing baseline problem, is a problem that structures most research on covert activities. This should not dissuade research on the issue, but rather means one has to situate one's claims about the public incidents, with the caveat of a potentially deviating non-observed baseline (i.e. no assumption of representativeness of the public cases is permissible). As explained above, however, we can infer that the majority of incidents are not publicly attributed from private threat reporting. The missing baseline problem is mitigated in the case of analysing public attribution, as the reasons to publicly attribute can often be established. The harder case, i.e. when not to publicly attribute, has yet to be robustly researched.

17 A whole host of private sector produced reports on state-sponsored cyber campaigns document this claim; for example, Novetta. Operation SMN: Axiom Threat Actor Group Report. Novetta, McLean, VA, 2014. https://perma.cc/53VG-SU7P.

18 The Five-Eyes consist out of the USA, UK, Canada, Australia and New Zealand.

19 On 15 May 2017, Neel Mehta, a Google security researcher, went public with a tweet indicating a similarity between WannaCry and a malware sample previously used by Lazarus, a North Korean advanced persistent threat actor. Neel Mehta used the hashtag #WannCryptAttribution, a very clear signal of what the tweet was intending to do.

20 Part of this analysis on NotPetya has been previously published in [45].

21 New Zealand did not independently assess it, but joined the Five Eyes in the condemnation, whilst the Canada attributed NotPetya to actors in Russia.

the trust in just one government. Thus, attributing in a coalition can be read as a credibility enhancing practice.

Whether the norm shaping effect is attained will greatly depend on if the audience buys the message, i.e., whether the potential pool of state adversaries actually care about being publicly outed, and thus accept the proposed interaction order. Differential impacts are likely. For some states, including Russia and North Korea, being outed publicly may have a neutral to slightly positive effect, bolstering their reputation as powers with both the ability and willingness to engage in effects operations to attain their strategic aims. For those particular countries, the necessity to take measures beyond the public attribution statements to impose costs may be applicable. However, the target audience of public attribution statements is broader than the specific adversary blamed. Some countries will care a lot about the risk of being outed publicly. For them, the risk of being publicly shamed by a large group of states may lead to a more careful tailoring of their offensive operations. Particularly countries still building up their offensive capabilities (see [89]) may adapt their policies and procedures to prevent indiscriminately delivering effects, as the risk of disclosure, when violating that imposed norm (all other things being equal), has increased. The norm would thus have a shaping effect. In that sense, if successful, the normative shaping of the operational space can have an independent deterrent effect on this subset of actors. The threat of collective public attribution can thus be read as a deterrence practice itself. Yet other countries will share the norm (through suasion, i.e. a logic of appropriateness) and join the effort of implementing it by ensuring their operations are aligned with the norm. They may even go as far as issuing public statements themselves.

As mentioned in the introduction, empirical examples where an incident is not publicly attributed for fear of undermining the interaction order would increase our confidence in public attribution serving a norm-shaping function. The author is aware of one very prominent case where that played a role: the hack against the Office of Personnel Management in the USA, where millions of security clearance applications were stolen. The Director of National Intelligence at the time, James Clapper, said about it in a Q&A session: 'You know on the one hand, please don't take this wrong way, you have to, kind of, salute the Chinese for what they did. If we had the opportunity to do that, I don't think we'd hesitate for a minute' [90]. This was followed by and immediate back-walking of the attribution statement, saying that China was only the leading suspect, not named as the perpetrator. He never confirmed the China attribution in official hearings. Rather, he contextualized his statement in his autobiography later: 'I hadn't thought the naming and shaming China would be the controversial part of my comments. To me, the important point – which wasn't quickly picked up on – was that I'd said China had hurt us dearly, but that it hadn't done anything outside the bounds of what nation-states do when conducting espionage' [90]. He then warned that the response would set a precedent that potentially would come back to haunt the USA when a similar collection opportunity arises. The OPM case strengthens our confidence that, at least in the case of the US policy on attribution, the matter of shaping the operational space looms large in the decision to attribute.

Alternative or joint explanations for public attribution

However, public attribution can serve further purposes beyond influencing the operational environment of adversaries. It is pertinent to lay them out here, as, on a case by case basis, they could serve as alternative or joint explanations for the public attribution of individual cyber intrusions. Because cyber security policymaking has shifted from a separate technical discussion into the 'normal' national security policymaking, the literature on state decision-making can partially explain public attribution. First, there are a whole host of potential domestic political motivations. They range from shifting the public's attention away from the government's inability to defend to classic bureaucratic politics, for example, by an agency wanting to get a modern public reputation as being competent in the cyber area. Public attribution also serves as a tool to create a baseline of activity with which one's own population judges the state of conflict between the home state and the adversary, potentially leading to a tying-hands effect, i.e. narrowing the policy space for domestic political actors, by fuelling demands for retaliation [91, 92].

Besides shaping the normative environment there are also other international motivations a state might have. A state may want to display its capability to attribute to raise the expected likelihood of being caught when targeting that state [93]. This is particularly relevant, as the perpetrators face an uncertainty distribution around the probability of being caught. Other states may join a public attribution campaign for diplomatic reasons, e.g., to use it as a demonstration of one's commitment to one's allies. Another strategic aim may be to build a community of defenders that cooperates in responding to cyber intrusions perpetrated by specific adversaries. Public attribution is one of the means of such coalition building, shining a light onto specific threat actors, and gaining momentum around defending against those. Soliciting public attribution from multiple states to gain political momentum can thus also be categorized as a strategic threat construction activity. Finally, public attribution can also be a way to directly engage an adversary, and, depending on the level of details released, to increase the global defensive community's ability to identify and defend against the particular adversary. The burning of adversary tooling, e.g. through submitting it to the public on VirusTotal, could be read as such an activity deliberately targeted at a particular adversary. Thus, public attribution can also be read as a counter-threat activity, i.e. keeping the adversary busy retooling, which could be gruelling and costly.

The years of experience with public attribution since Rid and Buchanan's seminal article have shown that, contrary to their suggestions, specificity in the data presented to support a public attribution claim is not a necessity to advance such a claim successfully (see further [47]). In the examples raised in this article (Bundestag, DNC hack, WannaCry, and NotPetya), the government initially did not readily offer up specific data to back up their attribution claims.[22] Rather, they built on the private sector claims, and used the government's reputation of possessing capable (signals) intelligence capabilities to convince the audience of the trustworthiness of their claims (see further [94]). In addition, in the two latter cases, as argued above, attribution coalitions were used to overcome the credibility problem. Two rationales may inform the lack of detail. First, from an intelligence perspective, the sources and methods informing the judgment may have been particularly sensitive. Second, the lack of detail may be motivated by the fact that technical evidence never 'proves' responsibility and is usually open to interpretation. By not offering any data, but using strong estimative language such as 'almost certain' or 'highly likely', the attributing entities were wagering their reputation as competent entities making attribution

22 In the case of WannaCry, the US government issued a very detailed criminal complaint against a North Korean operative, though with a delay of 9 months between the public claim and the publicizing of the court case. See *United States of America V. Park Jin Hyok*, MJ18-1479—Criminal Complaint (2018).

judgements (on what happens after public attribution see [46]). Other governments, e.g. the Swiss, had chosen yet an opposite route, publicizing detailed technical reports, but not publicly blaming a particular actor for the intrusion [95]. Further research should investigate the parameters shaping the different choices of whether and how to publicly attribute.

Conclusion

Attribution remains fundamental to the debate on how to respond to cyber intrusions. This article conceptualized public attribution by introducing two concepts, sense-making and meaning-making. Sense-making denotes the processes that help to establish an entity what happened. Meaning-making denotes the activities undertaken to reduce the uncertainty about an intrusion in targeted audiences and achieve political effects.

Building on this conceptual clarification, the article makes three main claims. First, we can understand public attribution better if we theoretically embed it in the literature explaining the revelation of covert activities, particularly Carson's work on interaction orders, and amend its conceptualization to actions in cyberspace. This adds clarity of what function public attribution can serve: i.e. signalling the rules of the 'game' (interaction order) with the intention to deter camera shy countries from engaging in a particular kind of covert activity, namely the one that is outside of those rules, and convincing allies of the merits of such an interaction order.

Second, in the cases investigated, contra the literature positing the 'attribution problem' as a formidable strategic challenge, the sense-making process of attribution was not the main policy challenge, but rather the policy response of what to do about the cyber intrusion, including whether to attribute publicly. The discussion of sense-making emphasized its generic quality comparing it to the domain of chemical warfare. Two empirical examples, the Bundestag and DNC incidents, illustrated how sense-making in the cyber domain can take place relatively rapidly.

Third, conceptualizing public attribution as a meaning-making strategy for attaining specific political effects is useful to understand its functions not currently captured in the literature. Public attribution is a particular meaning-making process, in which the attribution judgments are communicated publicly in order to reduce uncertainty in target audiences and attain political effects. Thereby, some intrusions are no longer treated on an incident basis, but politically are captured as campaigns of activity serving a particular strategic intent. Consequently, the responses are tailored accordingly, leading some states to engage in internationally coordinated campaigns of public attribution aimed at creating and sustaining a particular operational and normative environment, in which certain types of cyber operations are tolerated and others discouraged. The WannaCry and NotPetya attributions were used to illustrate three logics operative in public attribution: the creation of effects beyond the dyadic attacker–victim relationship, the shaping of the operational and normative space, and the limited deterrent value created through such actions. Thereby, the discussion was brought back to the tacit collusion framework, in particular, to the function of public attribution in signalling the rules of the 'game'. The argument is that in the long term, such multilateral public attribution has the potential to attain independent deterrent effects onto a specific sub-group of adversaries, particularly the ones who care about not being exposed.

There are limitations to the arguments made in this article. It cannot answer all the questions raised by the practice of public attribution. Particularly, it leaves out the cases of a government deliberately misattributing publicly, or representing a low confidence judgement as a certainty. Rather, it has focused on cases, in which a government gained a medium-to-high confidence judgement in the provenance of the cyber intrusions. In such cases, the question of whether and why to attribute publicly looms large. In addition, the increased ability to attribute is temporally bounded as it stands in relation to advances in offensive tradecraft and operational security. Thus, attribution capabilities may degrade, due to the adversary's advancement in both. Such advancement can be anticipated. For example, possibly in response to advancement in code similarity analysis, the industry already witnesses adversaries shifting towards using more generic, open-source tooling [96]. In the near future, adversaries may also integrate the skilful use of artificial intelligence in offensive operations, or cloud attribution judgements by attempting more false-flag operations [96, 97]. For advanced adversaries, locking down information accesses (human and technical) to ones' own organizational processes by foreign intelligence organizations may also temporarily decrease the ability to attribute to the actual sponsor. Finally, this article could not yet establish why specific intrusions are attributed publicly and others not.

These limitations, however, open up at least three new focuses for research on public attribution. A first research focus should be on empirically documenting some of the decision-making processes leading towards or against public attribution. Comparative aspects are of interest. For example, why is it that France was extremely hesitant in publicly attributing 2015 TV5 Monde intrusions, whilst Germany attributed the 2015 Bundestag intrusions? Or why is it that Denmark and Japan joined the WannaCry attribution campaign, but Germany and France did not? A second research focus should expand the nuance between the different types of attributions. This could focus on differentiating who is attributing (e.g. head of state, foreign minister, agency, court, etc.), what type of incident it is (e.g. espionage, sabotage, subversion), what triggered the attribution (e.g. executive decision, public pressure, agency leak, etc.), and how much of a strategic policy making capacity does the attributing state have? In this type of analysis, one could also try to corroborate the government's public claims independently. Finally, a third research focus results from the insight that at the strategic level attribution processes are not unique to events in the cyber domain. Consequently, subsequent research should further integrate and apply the existing international relations literature with regard to state motivations to lay public blame onto specific actors. Such an effort could further normalize the integration of cyber interactions into the broader analysis of conflict and competition and shed light onto the domestic and international pressures guiding such actions.

Public attribution has arrived on the centre-stage of strategic interaction in cyber (in-)security. The use of public attribution as a means of statecraft in national security policy is here to stay. It will become even further normalized and differentiated. On the question of how states will publicly attribute, as the policy capacity of more states develops in this area, one can expect to see the fine-tuning of the specificity of the actor blamed (e.g. state, agency, sub-group, human being) and whether details are offered to bolster their conclusions. Furthermore, we may see an increase in the building of international coalitions to attribute collectively to specific threat actors.

This article argued that for some states the strategic challenge is no longer only about attributing cyber intrusions. It is about 'publicly' attributing cyber intrusions. The process of public attribution is one element affecting the larger framework of relations between states in the domain of offensive cyber operations, here referred to as interaction order or rules of the 'game'. Overall, the analysis has shown the importance of the meaning-making process to

understanding the politics of attribution and the rewards of closely examining its linkage to the politics of secrecy and exposure of covert activities of states.

Acknowledgements

Earlier versions of this article were presented at ISA 2018, ECPR 2018, ETH CSS Cyber Conference 2018, BSidesZH 2018, University of Oxford CTGA 2019, SAS 2019, EUISS conference, and at the Transatlantic Dialogue on Military Cyber Operations. I would like thank the participants and organizers at these events, the three anonymous reviewers, as well as my colleagues at the Center for Security Studies (ETH Zurich) and at the Centre for Technology and Global Affairs (University of Oxford) for their insightful feedback. Thanks also to Jasper Frei for his excellent research assistance.

References

1. Stoll C. Stalking the Wily Hacker. *Commun ACM* 1988;**31**:484–97.
2. Clark DD, Landau S. Untangling attribution. In: Committee on Deterring Cyberattacks (ed.), *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy*. Washington, DC: The National Academies Press, 2010, 25–40.
3. Boebert E. A survey of challenges in attribution. In: Committee on Deterring Cyberattacks (ed.), *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy*. Washington, DC: The National Academies Press, 2010, 41–52.
4. Morgan PM. Applicability of traditional deterrence concepts and theory to the cyber realm. In: Committee on Deterring Cyberattacks (ed.), *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy*. Washington, DC: The National Academies Press, 2010, 55–76.
5. Liff AP. Cyberwar: a new 'Absolute Weapon'? The proliferation of cyberwarfare capabilities and interstate war. *J Strat Stud* 2012;**35**:401–28.
6. Kello L. The meaning of the cyber revolution: perils to theory and statecraft. *Int Security* 2013;**38**:7–40.
7. Rid T, Buchanan B. Attributing cyber attacks. *J Strat Stud* 2015;**38**:4–37.
8. Lindsay JR. Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. *J Cybersecurity* 2015;**1**:53–67.
9. Lupovici A. The 'Attribution Problem' and the social construction of 'Violence': taking cyber deterrence literature a step forward. *Int Stud Perspect* 2016;**17**:322–42.
10. Tor U. 'Cumulative Deterrence' as a new paradigm for cyber deterrence. *J Strat Stud* 2017;**40**:92–117.
11. Nye JS. Deterrence and dissuasion in cyberspace. *Int Security* 2017;**41**:44–71.
12. Steffens T. *Auf Der Spur Der Hacker Wie Man Die Täter Hinter Der Computer-Spionage Enttarnt*. Berlin: Springer Vieweg, 2018.
13. Wendt A. On constitution and causation in international relations. *Rev Int Stud* 1998;**24**:101–18.
14. Dray W. 'Explaining What' in History. In: Gardiner Patrick L. (ed.), *Theories of History*. London: Collier Macmillan, 1959.
15. Guitton C. Achieving attribution. Ph.D. Thesis. King's College London, 2014.
16. Lin H. Attribution of malicious cyber incidents: from soup to nuts. *J Int Affairs* 2016;**70**:75–106.
17. US Office of the Director of National Intelligence: A Guide to Cyber Attribution, NIC 1805-00278, 14 September 2018, https://perma.cc/7AJJ-P2UD.
18. The Grugq. "Security, Cyber, and Elections (part 2)", Medium (12 November 2016), https://perma.cc/NDT4-EVKE.
19. Boin A, t' Hart P, Stern E, *et al*. *The Politics of Crisis Management: Public Leadership under Pressure*. Cambridge: Cambridge University Press, 2005.
20. Egloff FJ. Cybersecurity and non-state actors: a historical analogy with mercantile companies, privateers, and pirates. DPhil Thesis. University of Oxford, 2018.
21. Carson A. Facing off and saving face: covert intervention and escalation management in the Korean War. *Int Organ* 2016;**70**:103–131.
22. Carson A. *Secret Wars: Covert Conflict in International Politics*. Princeton: Princeton University Press, 2018.
23. Borghard ED, Lonergan SW. Cyber Operations as Imperfect Tools of Escalation. Strategic Stud Q. 2019;**13**:122–145.
24. Valeriano B, Jensen BM, Maness RC. *Cyber Strategy: The Evolving Character of Power and Coercion*. New York: Oxford University Press, 2018.
25. Fischerkeller MP, Harknett RJ. Deterrence is not a credible strategy for cyberspace. *Orbis* 2017;**61**:381–93.
26. Aldrich RJ. *GCHQ: The Uncensored Story of Britain's Most Secret Intelligence Agency*. London: HarperPress, 2011.
27. Bamford J. *The Shadow Factory: The Ultra-Secret NSA from 9/11 to the Eavesdropping on America*. 1st edn. New York: Doubleday, 2009.
28. Warner M. Cybersecurity: a pre-history. *Intell Nat Security* 2012;**27**:781–99.
29. Lindsay JR. Cyber Espionage. In: Cornish Paul (ed.), *The Oxford Handbook of Cybersecurity*. Oxford: Oxford University Press, forthcoming.
30. Rovner J. Cyber war as an intelligence contest. *War on the Rocks*, 16 September 2019, https://perma.cc/3VD5-ARRH.
31. Guerrero-Saade JA. The ethics and perils of APT research: an unexpected transition into intelligence brokerage. In: *Virus Bulletin Conference*, 2015. Prague, Czech Republic.
32. Libicki M. The coming of cyber espionage norms. *Paper presented at the 9th International Conference on Cyber Conflict (CyCon)*, 30 May–2 June 2017.
33. Boeke S, Broeders D. The demilitarisation of cyber conflict. *Survival* 2018;**60**:73–90.
34. Georgieva I. The unexpected norm-setters: intelligence agencies in cyberspace. *Contemp Security Policy* 2020;**41**:33–54.
35. Pouliot V. The logic of practicality: a theory of practice of security communities. *Int Organ* 2008;**62**:257–88.
36. Finnemore M, Hollis DB. Beyond naming and shaming: accusations and international law in cybersecurity. *Temple University Legal Studies Research Paper* 2019;**14**:1–31.
37. Broeders D, Boeke S, Georgieva I. Foreign intelligence in the digital age. Navigating a state of 'unpeace'. The Hague Program for Cyber Norms Policy Brief. September 2019.
38. Buchan R. *Cyber Espionage and International Law*. Oxford, UK: Hart, 2019.
39. Carnegie A, Carson A. The spotlight's harsh glare: rethinking publicity and international order. *Int Organ* 2018;**72**:627–57.
40. Fischerkeller MP, Harknett RJ. Persistent engagement, agreed competition, cyberspace interaction dynamics and escalation, institute for defense analysis (2018), https://perma.cc/SQK3-Q3UL.
41. Carnegie A, Carson A. The disclosure dilemma: nuclear intelligence and international organizations. *Am J Polit Sci* 2019;**63**:269–85.
42. Poznansky M, Perkoski E. Rethinking secrecy in cyberspace: the politics of voluntary attribution. *J Global Security Stud* 2018;**3**:402–16.
43. Cormac R, Aldrich RJ. Grey is the new black: covert action and implausible deniability. *Int Affairs* 2018;**94**:477–94.
44. Deibert RJ. Toward a human-centric approach to cybersecurity. *Ethics & Int Affairs* 2018;**32**:411–24.
45. Eichensehr KE. Decentralized cyberattack attribution. *AJIL Unbound* 2019;**113**:213–17.
46. Egloff FJ, Wenger A. Public attribution of cyber incidents. *CSS Anal Security Policy* 2019;**244**:1–4.
47. Egloff FJ. Contested public attributions of cyber incidents and the role of academia. *Contemp Security Policy* 2020;**41**:55–81.
48. Grindal K, Kuerbis B, Badiei F, *et al*. Is it time to institutionalize cyberattribution? https://perma.cc/XN46-CWC2.
49. Satter R. The notification gap, *Medium*, 26 November 2017, https://perma.cc/5LA9-MK7Q.
50. Berghel H. On the problem of (cyber) attribution. *Computer* 2017;**50**:84–89.
51. Guerrero-Saade JA, Raiu C. Walking in your enemy's shadow: when fourth-party collection becomes attribution hell. In: *Virus Bulletin Conference*, 2017. Madrid, Spain.

52. Murphy EE. *Inside the Cell: The Dark Side of Forensic DNA*. New York: Nation Books, 2015.
53. Annex on Implementation and Verification, *Convention on the Prohibition of the Development, Production, Stockpiling, and Use of Chemical Weapons and on Their Destruction*, §26 Part XI(D).
54. OPCW Conference of the States Parties, *Addressing the Threat from Chemical Weapons Use, Decision Document, C-SS-4/DEC.3*, 27 June 2018, particularly §19.-21.
55. OPCW Conference of the States Parties, *Report of the Fourth Special Session of the Conference of the States Parties, C-SS-4/3*, 27 June 2018, §3.15.
56. OPCW-UN Joint Investigation Mechanism, *Third report of the OPCW-UN Joint Investigative Mechanism, S/2016/738/Rev.1*, 24 August 2016, Annex I.
57. Bundestags-Hack: Merkel Und Der Schicke Bär. *Die Zeit* (10 May 2017).
58. Digital Attack on German Parliament: Investigative Report on the Hack of the Left Party Infrastructure in Bundestag. https://perma.cc/58C4-DCV7.
59. Wir veröffentlichen Dokumente zum Bundestagshack: Wie man die Abgeordneten im Unklaren ließ. https://perma.cc/7CGS-SMWK.
60. Hacker-Angriff Auf Den Bundestag: Generalbundesanwalt Übernimmt Ermittlungen. *Spiegel* (20 January 2016), https://perma.cc/ZAH2-8MDQ.
61. Verfassungsschutz Warnt Vor Attacken Aus Russland. *Spiegel* (13 May 2015), https://perma.cc/H46A-7RJJ.
62. Merkel A. *Deutscher Bundestag*, 19. Wahlperiode, 159. Sitzung, 13 May 2020, p. 19700.
63. All Signs Point to Russia Being Behind the DNC Hack. *Motherboard* (25 July 2016), https://perma.cc/L322-ZWWK.
64. Rid T. Disinformation: A Primer in Russian Active Measures and Influence Campaigns. *Hearing in Front of the United States Senate Select Committee on Intelligence 115th Cong*ress, 2017. https://perma.cc/3FA6-22XW.
65. How Russia pulled off the biggest election hack in U.S. history. *Esquire* (20 October 2016), https://perma.cc/SCQ4-PL67.
66. Gioe DV. Cyber operations and useful fools: the approach of russian hybrid intelligence. *Intell Natl Security* 2018;**33**:954–20.
67. Obama's secret struggle to punish Russia for Putin's election assault. *Washington Post* (23 June 2017), https://perma.cc/PAS9-RBS8.
68. Feinstein, Schiff Statement on Russian Hacking. https://perma.cc/DF4G-44EN.
69. Joint Statement from the Department of Homeland Security and Office of the Director of National Intelligence on Election Security. https://perma.cc/L5UT-QX6X.
70. National Intelligence Council (ODNI), Background to 'Assessing Russian Activities and Intentions in Recent US Elections'. The Analytic Process and Cyber Incident Attribution. U.S. Office of the Director of National Intelligence, Washington DC, 2017.
71. Cavelty MD, Egloff FJ. The politics of cybersecurity: balancing different roles of the state. *St Antony's Int Rev* 2019;**15**:37–57.
72. U.S. Office of the Inspector General. A Review of Various Actions by the Federal Bureau of Investigation and Department of Justice in Advance of the 2016 Election, *U.S. Department of Justice*, 2018, https://perma.cc/8EVS-P9BF.
73. An Example of Deterrence in Cyberspace. https://perma.cc/EM3V-NDNQ.
74. Not the Cyber Deterrence the United States Wants. *Net Politics* (11 June 2018), https://perma.cc/ACE9-3W86.
75. Sanger DE. *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age*. New York: Crown, 2018.
76. FSB Finds Cyber-Spying Virus in Computer Networks of 20 State Authorities. *TASS* (30 July 2016), https://perma.cc/5C6Q-5K3X.
77. Buchanan B. *The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations*. Oxford: Oxford University Press, 2017.
78. Comptroller and Auditor General. Investigation: WannaCry Cyber Attack and the NHS. London: National Audit Office, 2017. https://perma.cc/KQ8M-CR6J.
79. Mehta N. Tweet: 9c7c7149387a1c79679a87dd1ba755bc @ 0x402560, 0x40f598 Ac21c8ad899727137c4b94458d7aa8d8 @ 0x10004ba0, 0x10012aa4 #WannaCryptAttribution, *Twitter*, 15 May 2017, https://perma.cc/T7SQ-MEJV.
80. Wright J. Cyber and International Law in the 21st Century, *Chatham House Royal Institute of International Affairs*, 23 May 2018.
81. Office of the Coordinator for Cyber Issues. Recommendations to the President on Deterring Adversaries and Better Protecting the American People from Cyber Threats. *U.S. Department of State*, 31 May 2018, https://perma.cc/A8UG-QEXN.
82. Egan BJ. International law and stability in cyberspace. *Berkeley J Int Law* 2017;**35**:169–80.
83. United States Government—White House. *National Cyber Strategy of the United States of America*. Washington DC, 2018.
84. United Kingdom of Great Britain and Northern Ireland, Foreign and Commonwealth Office. *UK's approach to the attribution of cyber incidents*. Personal copy.
85. Finnemore M, Hollis DB. Constructing norms for global cybersecurity. *Am J Int Law* 2016;**110**:425–79.
86. Finnemore M, Sikkink K. International norm dynamics and political change. *Int Organ* 1998;**52**:887–917.
87. Eichensehr KE. The law & politics of cyberattack attribution. *UCLA Law Rev* **67**: forthcoming.
88. Roguski P. Russian cyber attacks against georgia, public attributions and sovereignty in cyberspace. *Just Security*, 6 March 2020. https://perma.cc/99H3-KFGL.
89. Smeets M. Going cyber: the dynamics of cyber proliferation and international security. DPhil Thesis. University of Oxford, 2017.
90. Clapper JR, Brown T. *Facts and Fears: Hard Truths from a Life in Intelligence*. New York: Viking, 2018.
91. Giles K, Hartmann K. 'Silent Battle' Goes Loud: Entering a New Era of State-Avowed Cyber Conflict. *Paper presented at the 11th International Conference on Cyber Conflict (CyCon)*, 28–31 May 2019.
92. Fearon JD. Signaling foreign policy interests: tying hands versus sinking costs. *J Conflict Res* 1997;**41**:68–90.
93. Baram G, Sommer U. Covert or not covert: national strategies during cyber conflict. In: *11th International Conference on Cyber Conflict (CyCon)*, Tallinn, Estonia, 2019.
94. Grotto A. Deconstructing cyber attribution: a proposed framework and Lexicon. *IEEE Security & Privacy* 2020;**18**:12–20.
95. GovCERT.ch. APT Case Ruag—Technical Report, Bern: Government of Switzerland 2016. https://perma.cc/2XKP-4FAX.
96. Raiu C. Attribution 2.0. In: *area41 conference, Zurich*, Zurich, Switzerland, 2018. https://perma.cc/5FZ4-6SJR.
97. Bartholomew B, Guerrero-Saade JA. Wave your false flags! deception tactics muddying attribution in targeted attacks. In: *Virus Bulletin Conference, Denver CO*, 2016.