There Is No Attribution Problem, Only a Diplomatic One

https://www.e-ir.info/2020/03/22/there-is-no-attribution-problem-only-a-diplomatic-one/

EVA-NOUR REPUSSARD, MAR 22 2020

Both North Korea and Iran have been contentious actors in the cyberspace, being accused of several attacks in the past couple of years and both being listed on many occasions on CSIS's significant cyber incidents list (2019). North Korea has been accused of carrying out the Sony Pictures hack in 2014, the WannaCry Malware in 2017, several bank heists in Asia and cyberattacks on Indian nuclear plants in 2019 (Onaran, 2018; Findlay and White, 2019; White and Buseong, 2019). Iran, for its part, has been accused of attacking American banks in 2012 and Saudi infrastructures on multiple occasions, with the Shamoon virus in 2012 as well as the Aramco hack the same year, and more recently a cyberattack in a Saudi petrochemical company (Perlroth and Krauss, 2018). The 'attribution problem' in the cyberspace, i.e. the idea that identifying the source of a cyberattack is often complicated (Newman, 2016) is not supported by those examples, as Iran and North Korea have been identified by several countries as the main suspects of such attacks. Despite a lack of physical evidence behind those cyberattacks, cyber security firms and intelligence agencies governments have managed to attribute them to Iran and North Korea with high confidence.

Determining who is to blame in a cyberattack is a significant challenge, as the cyberspace is intrinsically different from the kinetic one: there is no physical activity to observe and technological advancements have allowed perpetrators to be harder to track and to remain seemingly anonymous (Brantly, 2016). Intelligence agencies have, however, often managed to identify suspects, not mainly through technical ways, but rather through clandestine and diplomatic ways (Rid, 2013). To achieve attribution, it is not enough to identify the suspects, i.e. the actual persons involved in the cyberattacks. Intelligence agencies must then determine if the cyberattacks had a political motive and were supported by a government or a non-state actor, with enough evidence to support diplomatic, military or legal options.

First, I will argue, as the North Korean and Iranian cases show, that attribution is possible in cyberspace and is far more a problem in theory than it is in practice. Cyberattacks occur in specific contexts and through technological, diplomatic and clandestine means, intelligence agencies do achieve high confidence attributions — seemingly rendering attribution no longer a problem in cyberspace. Attribution, however, remains inherently uncertain: the surge in false flag attacks, along with the difficulty in proving complicity between states and cyber aggressors lead to perpetrators to deny cyberattacks: that is the plausible deniability policy. I will, then, argue that plausible deniability creates a deterrence failure in the cyberspace.

I: High Confidence Attribution Is Achievable

I will argue that the attribution problem for cyberattacks is harder to solve in theory than it is in practice. Technological evidence, along with diplomatic knowledge and intelligence insight, has allowed for high confidence attributions to be made — seemingly solving the attribution problem.

The need for attribution is not limited to the cyberspace and is necessary for states' security: to retaliate, states' ought to know who their enemy is. Cyberattacks are, however, intrinsically different from conventional attacks, as there is no physical activity to observe, and professional attackers can sometimes manage to keep their anonymity

and hide their attacks, with cyberattack becoming more sophisticated (Lindsay, 2015; Lin, 2016). Before the attribution problem, it is important to identify that an attack is occurring, what the target is and how to protect it effectively (Brantly, 2016). Once it has been established, and whether states managed (or not) to protect their assets, the attribution problem arises — i.e., knowing who conducted the attack. Cyberattacks do leave technical clues that can be analysed: they can involve IP addresses, log file analysis, languages of the code, code similarity with previous malware… Those clues are, however, predominantly circumstantial and are not enough to attribute a cyberattack with high confidence (Hunker et al., 2008). From a solely technical perspective, Bartholomew and Guerrero-Saade from Kaspersky Lab do argue that accurate attribution is almost impossible, notably because of the existence and surge of false flag attacks (2016). False flag attacks can lead to misattribution which may undermine a state's credibility and lead to political differences in the event of economic or diplomatic sanctions. In cyberspace, false flag attacks are easier to carry out than in the physical space, as attackers can more easily deceive their origins. They can notably use previously used code of other groups — as it has been the case in the Pyeongchang hack, which Russian hackers tried to blame on North Koreans. Despite the seeming difficulty of attributing cyberattacks, the Pyeongchang hack has been attributed with high confidence to Russia. As Goodman argues, attribution is easier in practice than it is in theory (2010): whether it is the Iranian and North Korean cases cited earlier in this essay, or the several cyberattacks attributed to Russia, such as NotPetya, the December 2015 Ukrainian power grid hack, and the DNC hack— high confidence attribution is empirically possible in cyberspace.

High confidence attribution is, indeed, possible, because security analysts and intelligence firms do not solely rely on technical clues to find the perpetrators. Attribution is a multidimensional problem that relies on all sources of information available to be solved: forensics, human intelligence, signals intelligence, history and geopolitics (Bartholomew and Guerrero-Saade, 2016; Lin, 2016). Using the example of the US, Kugler argues that considerable cyberattacks are likely to arise out of a specific strategic context and with a specific political motive, which makes possible to attribute to specific attackers (2009). While cyberattacks occur in specific contexts and are often carried out by opponents ahead or along with kinetic warfare (ODNI, 2018) — states must be wary and not hasten the attribution process. It is a lengthy process that requires more than motive and circumstantial evidence, as the Winter Olympics in Pyeongchang hack perfectly illustrated. Technical clues and political motive first led cyberanalysts to believe that it was a North Korean hack. The FBI had, however, some intelligence insights which allowed for the hack to be attributed with high confidence to Russia (Greenberg, 2018). In the case of the Sony Pictures hack, similarly, it was intelligence insights along with political motive, technological clues and human errors — that have allowed the FBI to attribute the cyberattack to the North Korean government (FBI, 2014; VanDerWerff and Lee, 2015). As Rid and Buchanan (2015) argue, the quality of attribution is likely to improve as the number of intelligence sources increase. Earlier, it was mentioned that cyberattacks were becoming more sophisticated, which seemingly makes it harder for cyber security firms and intelligence agencies to attribute the attack with high confidence. This sophistication, however, is in itself a clue on who the perpetrators might be: sophisticated cyberattacks are only capable by a few states, drastically diminishing the possible suspects. Cybersecurity firms believed that only Iran, China, Russia, the United States and Israel had the technical sophistication to carry out the petrochemical attack in Saudi Arabia. The attack was immediately attributed to Iran, as China and Russia have energy deals with Saudi Arabia, while Israel and the United States have been Saudi Arabia's diplomatic allies against Iran (Perlroth and Kraus, 2018).

This section has aimed to argue that, contrary to what is often argued, the attribution problem is far much more a problem in theory than in practice. Along with intelligence and technical capabilities, cyberattacks always occur in a context, which allows intelligence agencies to attribute them with high certainty. High confidence attribution is possible and common, but *what next?*

II: …but Inherent Uncertainty Remains a Problem

Earlier in this essay, it was argued that through technological and clandestine means, cyberattacks could be attributed with high certainty. The following part of this essay moves on to describe, in greater detail, that the inherent uncertainty of attribution of cyberattacks and their nature allows for plausible deniability from suspected states. Those factors, along with the identity of the perpetrators, contribute to making deterrence by punishment harder for targeted states and cause cyberattacks to remain fairly inconsequential for the perpetrators.

Attribution remains uncertain because of states' plausible deniability: false flag attacks are becoming more common and the complicity between the perpetrators and a state is difficult to prove. Iran and North Korea, have generally denied their involvement in cyberattacks, despite consensuses across cyber security firms and intelligence agencies (Reuters; 2012; 2014; 2019). Iran might have been blameless in some of those accusations: in 2019, the UK National Cyber Security Centre and the NSA just revealed that a group of Russian hackers had carried out attacks in more than 35 countries and tried to blame Iran for it (Financial Times, 2019). The surge of false flags attacks and the deniability from the accused states are part of the reason why there still is an attribution problem in cyberspace (Bartholomew and Guerrero-Saade, 2016). It is not enough to determine the identity of the attackers and identifying false flag attacks: intelligence agencies need to identify the states or actors that sponsored the attacks (Pihelgas, 2015). Establishing attribution is one thing, but establishing complicity with a state is another completely. It is possible to track attackers and their geographic location, but it is more difficult to establish any formal government role in the cyberattack (Singer and Friedman, 2014).

Attempts at attributions are further complicated by the fact that hackers often deliberately downplay their affiliation with states and call themselves 'hacktivists'. In those cases, technical clues and motive could lead intelligence agencies to attribute attacks to certain states despite the latter being unaware of the cyberattack — i.e., it could lead to misattribution. Intelligence agencies have the heavy task of proving complicity beyond a reasonable doubt, between hackers and states (ODNI, 2018). In the case of WannaCry attacks, the US Department of Justice managed to assign the blame to a North Korean hacker, Park Jin Hyok; however, it failed to establish complicity with the North Korean state (Cimpanu, 2018). Similarly, in September 2019, the US Department of the Treasury imposed sanctions on three North Korean state-controlled hacking groups, notably the ones behind Sony Pictures hack, WannaCry and the bank heists against Asian backs (Cimpanu, 2019). In cyberspace, attribution can never be certain and accused states will argue that there is no complicity from their government with the attackers. This claim is hard to refute and requires prior intelligence to argue with certainty that the actors were complicit.

This inherent uncertainty and the nature of the attacks lead to the partial failure of deterrence in cyberspace – that is, discouraging other states and international actors from taking unwanted actions. As far as deterrence by denial is concerned, the US, in particular, has been successfully raising the costs and risks of an attack, and cyberattacks are, indeed, getting more sophisticated (Lindsay, 2015). Earlier in this essay, it was argued that the sophistication of cyberattacks meant that only a few states were capable of carrying them out, which means that deterrence by denial works. Deterrence by punishment, however, is failing in cyberspace as attacks remain fairly inconsequential for perpetrators. Attribution is only the first step in creating a system of deterrence by punishment, but ought to be followed by retaliation from targeted states (Hunker, et al., 2008). Intelligence insights do allow for high confidence attribution despite plausible deniability from sponsored states. High confidence attributions, however, often rely on classified information, which creates doubts in the attribution (Libicki, 2009; Tran, 2018). Doubtful attribution can prevent states from making needed policy decisions against the perpetrators, that would not be supported by the people or international actors. The importance of convincing third parties is essential if a state seeks to retaliate by any means, whether legal, diplomatic or military. In the case of the Sony Pictures hack, the FBI has been withholding the evidence that led them to accuse North Korea: it then became harder for the US government to take legal measures against the DPRK, as several experts claimed it was likely to have been carried out by 'hacktivists' rather than sponsored by the North Korean state (Goldsmith, 2014; Zetter, 2014).

Cyberattacks do support the claim that denial strategies are often more reliable than punishment strategies — as cyber aggressors are yet to be proportionally punished (Mazarr, 2018). Empirically, cyberattacks which have been attributed beyond a reasonable doubt, have largely remained unpunished for several reasons. First, in the case of an imbalance of power between states, the targeted state cannot retaliate against its enemy. Despite Estonia successfully attributing its cyberattacks in 2007 to Russia, it could not exploit this responsibility because of the geopolitical imbalance between the two states (Goodman, 2011). In other cases, cyberattacks are ongoing without the targeted state being aware of it. In the case of Stuxnet, for example, it took the Iranian government several years to realise it had been hacked in the first place, and that the problem did not come from an engineer (Zetter, 2014). In other cases, the nature of cyberattacks makes it harder for states to retaliate: if data has been stolen, it cannot be recovered. A well-conducted cyberattack, therefore, is an effective way of attacking a state, its important infrastructures (nuclear plants, electricity grids, etc.) without having to suffer any retaliation.

Cyberspace is intrinsically different from the kinetic space, and achieving proportionate retaliation is more complicated than in the case of conventional attacks. High confidence attributions will, therefore, compel targeted states to make a judgement call on how it should react, if it should retaliate, and how. In the several examples cited throughout this essay – North Korean, Iranian, and Russian— cyberattacks have remained fairly inconsequential for the perpetrators despite high confidence attribution. Those countries cannot suffer from the bad publicity emanating from the attribution (contrary to the US and Stuxnet) — and are already subject to a variety of economic sanctions ; therefore, they hardly can be deterred by the US. It is, therefore, unlikely that the situation changes and that North Korea or Iran stop carrying out cyberattacks — unless targeted states retaliate. Deterrence by punishment is likely to be more successful if cyberattacks triggered kinetic ones: targeted countries would, however, take the risk of escalating the conflict. That is why conventional retaliation is far more likely to happen if both states are already engaged in a conventional war, or if the attack is severe enough and causes substantial material or human damages, similar to a conventional attack (Tsagourias, 2012; Rid, 2013).

This section showed that once an international actor is a victim of a cyberattack, attributing a cyberattack is only the first step in the attribution problem. The cyber attribution problem then takes another form and becomes the cyber deterrence problem. As states' complicity is hard to prove, and proportional retaliation is complicated in the cyberspace — states usually fail to deter cyberattacks.

Conclusion

In the past years, many cyberattacks have been attributed with a high degree of confidence to North Korea and Iran. I aimed to show that the attribution problem in cyberspace is much more a problem in theory than it is in practice. In theory, cyberattacks can be carried out anonymously — in practice, however, geopolitical context, technological tools and intelligence insights allow to attribute the attacks. High confidence attribution is possible and common, seemingly solving the attribution problem.

I argued, however, that attribution remains a problem in cyberspace for one main reason: attribution is inherently uncertain as suspected states will always have sufficient deniability, whether it is because of false flag attacks, or by putting a distance between them and the attackers. This inherent uncertainty makes proportional retaliation harder for targeted states, as it will entail a judgement call. Cyberattacks highlight a deterrence failure in the cyberspace, as states who sponsor such attacks go largely unpunished, and are not discouraged in carrying out cyberattacks. Iran and North Korea can hardly be subject to more international sanctions or worst publicity — hence, they have no incentive in stopping cyberattacks. Only when a cyberattack will be severe enough, could targeted countries resort to military means, leaving one wondering whether cyberattacks could trigger a proportional kinetic attack that would be justified with regards to international law.

Bibliography

Brantly, A. F. (2016) 'Anonymity and Attribution in Cyberspace' in The Decision to Attack: Military and Intelligence Cyber Decision-Making. University of Georgia Press.

Bartholomew, B. And Guerrero-Saade, J. A. (2016) Wave Your False Flags! Deception Tactics Muddying Attribution in Targeted Attacks. Virus Bulletin Conference October 2016. Kaspersky Lab.

Bronk, C. and Tikk-Ringas, E. (2013) The Cyberattack on Saudi Aramco. Survival, 55:2. pp. 81-89.
Cormac, R. and Aldrich, R. J. (2018) Grey Is the New Black: Covert Action and Implausible Deniability. International Affairs, 94: 3. pp. 477-494

FBI (2014) Update on Sony Investigation. FBI [ONLINE] 19. December 2014. Available at: https://www.fbi.gov/news/pressrel/press-releases/update-on-sony-investigation [Accessed 5. November 2019]

Cimpanu, C. (2018) How US authorities tracked down the North Korean hacker behind WannaCry. ZDNet [ONLINE] 6. September 2018. Available at: https://www.zdnet.com/article/how-us-authorities-tracked-down-the-north-korean-

hacker-behind-wannacry/ [Accessed 2. December 2019]

Cimpanu, C. (2019) US Treasury sanctions three North Korean hacking groups. ZDNet [ONLINE] 13. September 2019. Available at: https://www.zdnet.com/article/us-treasury-sanctions-three-north-korean-hacking-groups/ [Accessed 1. December 2019]

CSIS (2019) Significant Cyber Incidents. Centre for Strategic and International Study [ONLINE] Available at: https://www.csis.org/programs/technology-policy-program/significant-cyber-incidents [Accessed 4. December 2019]

Financial Times (2019) Russian cyberattack unit 'masqueraded' as Iranian hackers, UK says. Financial Times [ONLINE] 21. October 2019. Available at: https://www.ft.com/content/b947b46a-f342-11e9-a79c-bc9acae3b654 [Accessed 10. December 2019]

Findlay, S. and White, E. (2019) India confirms cyberattack on nuclear power plant. Financial Times [ONLINE] 31. October 2019. Available at: https://www.ft.com/content/e43a5084-fbbb-11e9-a354-36acbbb0d9b6 [Accessed 5. November 2019]

Goodman, W. (2010) Cyber Deterrence: Tougher in Theory than in Practice? Strategic Studies Quarterly, 4:3. pp. 102-135

Goldsmith, J. (2014) The Sony Hack: Attribution Problems, and the Connection to Domestic Surveillance. Lawfare [ONLINE] 19. December 2014. Available at: https://www.lawfareblog.com/sony-hack-attribution-problems-and-connection-domestic-surveillance [Accessed 1. December 2019]

Goutam, R. J. (2015) The Problem of Attribution in Cyber Security. International Journal of Computer Applications, 131:7. pp. 34-36

Greenberg, A. (2018) Russian Hacker False Flags Work—Even After They're Exposed. WIRED [ONLINE] 27. February 2018. Available at: https://www.wired.com/story/russia-false-flag-hacks/ [Accessed 5. November 2019]

Hunker, J., Hutchinson, R. and Margulies, J. (2008) 'Attribution of Cyberattacks on Process Control Systems' in Papa, M. and Shenoi, S. (eds) IFIP International Federation for Information Processing, Volume 290. Boston: Springer. pp. 87–99

Section Author(s): Richard L. Kugler, R. L. (2009) 'Cyberpower and National Security' in Kramer, F. D., Starr, S. H. and Wentz, L. K. (eds) Deterrence of Cyberattacks. University of Nebraska Press, Potomac Books. pp. 309-340

Libicki, M. C. (2009) 'Why Cyberdeterrence Is Different' in Cyberdeterrence and Cyberwar. RAND Corporation.

Lin, H. (2016) Attribution of Malicious Cyber Incidents: From Soup to Nuts. Journal of International Affairs, 70:1. pp. 75-137

Lindsay, J. R. (2015) Tipping the Scales: the Attribution Problem and the Feasibility of Deterrence Against Cyberattack. Journal of Cybersecurity, 1:1. pp. 53–67

Mazarr, M. J. (2018) Understanding Deterrence. RAND Corporation.

Office of the Director of National Intelligence (2018) A Guide to Cyber Attribution. Office of the Director of National Intelligence [ONLINE] 14. September 2018. Available at: https://www.dni.gov/files/CTIIC/documents/ODNI_A_Guide_to_Cyber_Attribution.pdf [Accessed 25. November 2019]

Onaran, Y. (2018) North Korea Hackers Tried to Take $1.1 Billion in Bank Attacks. Bloomberg [ONLINE] 8. October

2018. Available at: https://www.bloomberg.com/news/articles/2018-10-08/north-korea-hackers-broke-into-banks-tried-to-take-1-1-billion [Accessed 25. November 2019]

Newman, L. H. (2016) Hacker Lexicon: What Is the Attribution Problem? WIRED [ONLINE] 24. December 2016. Available at: https://www.wired.com/2016/12/hacker-lexicon-attribution-problem/ [Accessed 26. November 2019]

Perlroth, N. and Krauss, C. (2018) A Cyberattack in Saudi Arabia Had a Deadly Goal. Experts Fear Another Try.

The New York Times [ONLINE] 15. March 2018. Available at: https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html [Accessed 25. November 2019]

Perlroth, N. and Krauss, C. (2018) A cyberattack in Saudi Arabia failed to cause carnage, but the next attempt could be deadly.

Independent [ONLINE] 21. March 2018. Available at: https://www.independent.co.uk/news/long_reads/cyber-warfare-saudi-arabia-petrochemical-security-america-a8258636.html [Accessed 25. November 2019]

Pihelgas, M. (2015) Mitigating Risks arising from False-Flag and No-Flag Cyberattacks. NATO Cooperative Cyber Defence Centre of Excellence.

Reuters (2012) From Sudan to cyber, secret war with Iran heats up. Reuters [ONLINE] 6. November 2012. Available at: https://www.reuters.com/article/us-iran-covert-war/from-sudan-to-cyber-secret-war-with-iran-heats-up-idUSBRE8A50G420121106 [Accessed 9. December 2012]

Reuters (2014) North Korea denies responsibility for Sony cyberattack. Reuters [ONLINE] 20. December 2014. Available at: https://www.reuters.com/article/us-sony-cybersecurity-northkorea-denial/north-korea-denies-responsibility-for-sony-cyber-attack-idUSKBN0JX24720141220 [Accessed 9. December 2019]

Reuters (2019) North Korea denies it amassed $2 billion through cyberattacks on banks. Reuters [ONLINE] 1. September 2019. Available at: https://www.reuters.com/article/us-northkorea-cyber/north-korea-denies-it-amassed-2-billion-through-cyberattacks-on-banks-idUSKCN1VM18K [Accessed 9. December 2019]

Rid, T. (2013) Cyber War Will Not Take Place. London: C. Hurst & Co.

Rid, T. and Buchanan, B. (2015) Attributing Cyberattacks. Journal of Strategic Studies, 38:1-2. pp. 4-37

Singer, P.W., and Friedman, A. (2014) 'Whodunit? The Problem of Attribution' in Cybersecurity and Cyberwar. Oxford University Press. pp. 72-76

Symantec (2018) The Cybersecurity Whodunnit: Challenges in Attribution of Targeted Attacks. Symantec [ONLINE] 3. October 2018. Available at: https://www.symantec.com/blogs/expert-perspectives/cyber-security-whodunnit-challenges-attribution-targeted-attacks [Accessed 25. November 2019]

Tran, D. (2018) The Law of Attribution: Rules for Attribution the Source of a Cyber-Attack. The Yale Journal of Law and Technology, 20:1. pp. 376-441

Tsagourias, N. (2012) Cyberattacks, Self-Defence and the Problem of Attribution. Journal of Conflict & Security Law, 17:2. pp. 229–244

VanDerWerff, E. T. and Timothy B. Lee, T. B. (2015) The 2014 Sony hacks, explained. VOX [ONLINE] 8. June 2015. Available at: https://www.vox.com/2015/1/20/18089084/sony-hack-north-korea [Accessed 25. November 2019]

Warrell, H. and Foy, H. (2019) Russian cyberattack unit 'masqueraded' as Iranian hackers, UK says. Financial Times [ONLINE] 24. October 2019. Available at: https://www.ft.com/content/19ba2210-f3fc-11e9-a79c-bc9acae3b654 [Accessed 7. November 2019]

White, E. and Buseong, K. (2019) Kim Jong Un's cyber army raises cash for North Korea. Financial Times [ONLINE] 18. June 2019. Available at: https://www.ft.com/content/cbb28ab8-8ce9-11e9-a24d-b42f641eca37 [Accessed 5. November 2019]

Zetter, K. (2014) Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon. Broadway Books.

Zetter, K. (2014) The Evidence That North Korea Hacked Sony Is Flimsy. WIRED [ONLINE] 17. December 2014. Available at: https://www.wired.com/2014/12/evidence-of-north-korea-hack-is-thin/ [Accessed 13. December 2019]

Written by: Eva-Nour Repussard
Written at: King's College University
Written for: Professor Joe Maiolo
Date written: December 2019