Does the delegation of cyber attacks to cyber proxies increase the risk of militarised interstate disputes due to information asymmetries?

ABSTRACT

The nature of cyberspace makes delegation of attacks from states to proxies increasingly attractive due to increases in plausible deniability. We first investigate the likely effects of this increase of uncertainty on militarized interstate disputes (MIDs) by surveying the literature on principal-agent theory and state-proxy relationships. We find that uncertainty likely encourages more proactive defence strategies, thereby increasing chances of conflict escalation. We posit that delegation of cyber attacks further exacerbates these effects, increasing the risk of MIDs, and propose a framework to empirically investigate our hypothesis and present future research avenues.

INTRODUCTION

Notable cyber breaches since 2007 triggered the enhancement of cybersecurity capabilities by states in response to potential threats (Liff, 2012). Yet strategic relations in cyberspace remain understudied. This paper analyzes the effect of the delegation of cyber attacks, by states to “cyber proxies”, on the risk of militarized interstate disputes. A better understanding of the implications of offensive cyber strategies can contribute to enhanced decision-making by states in cyberspace.

A Militarized Interstate Dispute (MID) is defined as “[a case] of conflict in which the threat, display or use of military force short of war by one member state is explicitly directed towards the government, official representatives, official forces, property, or territory of another state. Disputes are composed of incidents that range in intensity from threats to use force to actual combat short of war” (Jones *et al.,* 1996: 163).

Cyber proxies are non-state actors that conduct or directly contribute to an offensive cyber action that is enabled knowingly, whether actively or passively, to achieve political objectives on behalf of a patron state (Borghard and Lonergan, 2016; Maurer, 2018b). This definition is deliberately broad to encompass relationships where proxies are both tightly controlled as well as supported only indirectly by a government.

Maurer (2018b) suggests three types of proxy relationships: direct delegation, orchestration and permitting. *Direct delegation* implies the state has tight control over the agenda and actions of the proxy (Salehyan, 2010). *Orchestration* means proxies may receive tools or funding from the state, but no specific mandate (Maurer, 2018a); while *permitting* refers to the state indirectly benefiting from the proxy’s actions and turning a blind eye to their actions (ibid). Contrary to traditional alliances, states working with non-state actors in cyberspace prefer to work in secret (Borghard and Lonergan, 2016).

Cyberwarfare is conceptualised as a state of conflict with direct political or military objectives between two or more political actors characterized by computer network operations whose means are non-kinetic (Liff, 2012). Cyber attacks are a projection of power via computer network attacks (ibid). Cyber attacks may be part of a cyberwar, but can also stand alone to produce preferred outcomes within cyberspace or in the kinetic world (Maurer, 2018a).

Our argument is divided into four sections. The paper proceeds with a discussion of the literature on cyber proxies through the lens of the principal-agent framework and information asymmetries and uncertainty. Next, we present the main argument, our hypothesis and the underlying causal mechanism. The subsequent section proposes a quantitative research design to test the hypothesis. We conclude by summarising the main arguments, considering limitations and proposing future research avenues.

LITERATURE REVIEW

CYBER PROXIES

This section presents the state of the literature on the principal-agent theories before focusing on state-cyber proxy relationships.

The principal-agent framework describes how a relatively powerful actor delegates fighting to a subordinate actor to minimise the occurrence of a disturbance (Berman *et al.*, 2019). Delegation is likely when the expected costs of warfare for the principal are high, the tolerance for such costs is low and an agent can suppress disturbances at a relatively lower cost (ibid; Salehyan, 2010). When the principal and agent's interests are unaligned or when the principal lacks control over the agent, the agent might carry out actions that the principal did not intend, leading to agency loss (Maurer, 2018b). Delegation is, therefore, a successful strategy for the principal only if it increases the welfare of the principal counterfactually.

Analyses of the delegation of war using the principal-agent lens focus on foreign sponsorship of insurgent groups (Byman and Kreps, 2010; Salehyan, 2010; Salehyan *et al.*, 2014; Berkowitz, 2018). Recently, cyber warfare has been underscored as an ideal vehicle for proxy strategy. The importance of including non-state actors in the analysis of cyber conflict has been highlighted by Steiger *et al.* (2018), as non-state actors have developed capabilities conducive to states' political objectives. Case studies on the US, Russia, China, Iran and Syria provide insight into how states use cyber proxies to project power with varying principal-agent structures (Maurer, 2018a). Generally, acquiring strong cyber capabilities provides states with an advantageous asymmetric weapon due to the Internet's diffusion of reach (Maurer, 2018a); society's reliance on computers (Mumford, 2013); plausible deniability and offensive advantage in cyberspace (Liff, 2012). Cyberweapons act as a force-multiplier for conventional capabilities, making escalation cheaper (Smeets, 2018; Saltzman, 2013).

The appeal of conducting proxy wars in cyberspace boils down to the "alluring combination of plausible deniability and lower risk" (Mumford, 2013:45). Cyber proxies can disguise a state's identity to a greater extent than traditional proxies and information technology is easier to transfer than weapons (ibid).

Agency loss remains an issue in cyberspace, however. Proxies may be less reliable or controllable, but stronghold over a cyber proxy may undermine a state's desire to avoid attribution (Borghard and Lonergan, 2016).

UNCERTAINTY IN CYBERSPACE

This section assesses the literature on information asymmetries and conflict emergence, paired with the literature on how cyberspace exacerbates information asymmetries to present how states may act under uncertainty.

Given that war is always costly, states have incentives to locate settlements preferable to war (Fearon, 1995). When views about relative power (ibid) or the ability to absorb and inflict costs diverge (Reiter, 2003), this equilibrium is disrupted. Consequently, information asymmetries are an essential feature of the escalation of crises. Reed (2003) posits that uncertainty is a central cause of conflict: when a defender's decision to accept or reject any demand is a function of private information, the probability of war increases. States are more likely to miscalculate their bargaining leverage when unobservable factors such as resolve and military tactics prove consequential to victory (ibid). Powell (2004) draws attention to the mechanism through which actors convey information, which affects how long informational asymmetries take to resolve.

Uncertainty is a technical characteristic of cyberspace (Gomez, 2020). Scholars underscore technological vulnerabilities, potential for deception and increased dependence as reasons for such uncertainty (Dunn Cavelty, 2013; Forsyth and Pope, 2014; Gartzke and Lindsay, 2015). This is exacerbated by a lack of domain expertise, limiting decision-makers' ability to evaluate cyber attacks (Hansen and Nissenbaum, 2009). Even with complete information, decision-makers are conscious of possible misperception (Gomez, 2019; Borghard and Lonergan, 2017).

The attribution of cyber attacks is particularly problematic (Libicki, 2011; Farwell and Rohozinski, 2012; Lin, 2016; Lucas, 2017). A state can only be held accountable for the offensive actions of a cyber proxy "if that proxy is under tight control of a government and if the effect of the action causes significant harm" (Maurer, 2018a). Maurer (ibid) highlights that attribution is a question of degree. This is nuanced by Borghard and Lonergan (2016), who contend that attribution of the actor itself is not the main issue - cyber proxies are useful specifically because states *can* attribute attacks to the proxies. However, ascertaining the level of control over a proxy is less straightforward (ibid; Lin 2016). Even when states are technically capable of attribution, they may forgo doing so to avoid revealing capabilities to adversaries (Borghard and Lonergan, 2016). As stated by Libicki (2011), "proving that the other side may be vulnerable requires revealing the vulnerability."

There is little agreement in the literature on the occurrence of cyberwar (see e.g., Rid, 2012 and Liff, 2012 vs. Stone, 2013). Notwithstanding, information deficits contribute to the portrayal of cyberspace as an increasingly threatening domain and raises concerns about national security and how states should act in the face of cyber attacks (Gross *et al*., 2017). Traditional game theory states that the logic of the Prisoner's Dilemma drives two parties into conflict and that the lack of information and uncertainty over how an adversary will act increases adversarial strategies and mutual losses (Snyder, 1971). Some scholars postulate that states beset by uncertainty may increase technological capabilities (Libicki, 2011); favour offensive cyber operations (Jarvis *et al.*, 2017); or discourage adversaries with preemptive attacks (Saltzman, 2013; Smeets, 2018). This is reflected, for example, in the US Department of Defence (DoD) Cyber Strategy (2018:4), which calls for "defending forward to intercept and halt cyber threats and by strengthening the cybersecurity of systems and networks that support DoD missions".

To summarize, the possibility of cyberwar is introducing unprecedented levels of uncertainty to international relations, increasing chances of conflict escalation.

THEORETICAL ARGUMENT AND HYPOTHESIS

Our literature review exposes a lack of scholarship that assesses the effects of cyber proxy relationships on interstate conflict dynamics. This is pertinent to study because states are increasingly entering the cyber fray to pursue political objectives traditionally resolved through diplomacy, sanctions or kinetic force (Lucas, 2017; Weiss and Jankauskas, 2019). The implications of cyber proxy relationships matter for national security decision-making and international peace and stability. Thus, to investigate this relationship, our paper presents the following hypothesis:

H: the delegation of cyber attacks by states to cyber proxies increases the risk of MIDs.

The causal mechanism rests on a set of interlinked assumptions drawn from existing work. The starting assumption is that war is costly (Fearon, 1995). States thus wish to reduce the risk of MIDs. They do so by delegating cyber attacks to proxies to mask their identities. States may wish to carry out cyber attacks for myriad reasons, whether to protect economic prosperity, social stability or national security from threats (Hoffman, 2019). The literature suggests three main types of proxy relationships: direct delegation, orchestration and permitting (Maurer, 2018a). Any state may pursue any of these three relationships or "mix" them, though exhibit a significant degree of path dependence once a particular approach has been favoured (ibid).

Regardless of the proxy relationship a state chooses, states will tend to favour non-state proxies due to their ability to disguise a state's identity to a greater extent than other forms of proxies such as private military companies, adding a further layer of plausible deniability (Mumford, 2013). How states choose their cyber proxies - whether they are state-based hackers or groups based outside of nation-state borders - is beyond the scope of this paper. The key idea is that non-state actors have developed capabilities conducive to states' political objectives and are not bound by physical boundaries, hence these proxies can be based anywhere. For example, a cyber crime centre in Nairobi run by 77 Chinese nationals was discovered by Kenyan authorities in 2014 (*BBC News*, 2014). Other reports claim that Pyongyang's cyber attackers operate from India (Sanger *et al.,* 2017).

A myriad of literature concurs that the attribution of cyber attacks is difficult (Libicki, 2011; Farwell and Rohozinski, 2012; Lin, 2016; Lucas, 2017). The fact that cyber proxies can conduct cybercrime from outside the principal state's borders adds yet another layer of complexity to the problem of attribution. Even if an attack is attributed to the proxy, the level of control a state has over a proxy is difficult to ascertain (Borghard and Lonergan, 2016). As such, there is heightened ambiguity as the victim state has little to no verifiable information about the origin of the attack.

The victim state has two options: it can assume that the attack was delegated; or it can assume that the attack was not delegated. Regardless, the cyber attack must originate from a physical location within a state. Thus, either a state intended to harm the victim state through a cyber proxy, or a cyber proxy was acting independently, but it is nonetheless the host state's responsibility to maintain control of crime originating from within its borders. Given the uncertainty of the attacker, the victim state may respond by increasing its cyber capabilities (Libicki, 2011), not only against the suspected state, but against as many potential attacking states as possible. Assuming that outside states can perceive increases in the victim state's cyber capabilities and that signals in cyberspace are prone to misperception (Borghard and Lonergan, 2017), a security dilemma is likely to ensue.

The magnitude of the security dilemma depends on the offense-defense balance. When the offense has the upper hand, a state's reaction to international tension will increase the chances of war as incentives to preempt increase (ibid). Most scholars claim that cyberspace favours the offence (Slayton, 2016:72). Thus, outside states will respond to the victim state increasing its capabilities by also increasing their cyber capabilities because of uncertainty about the victim state's intentions. A spiral results by which the risk of MIDs increases not just between the original attacking-victim dyad, but among all dyads. Here, we are assuming a decentralised network of states that interact equally; even if alliances exist between certain actors, the responding increases in capabilities would occur among groups of states instead of individual states.

To summarise, states delegate cyber attacks to cyber proxies because they wish to reduce the risk of MIDs, yet this delegation likely culminates in the opposite effect. The following section proposes a quantitative research design to empirically test our hypothesis.

METHODOLOGY

The literature is dominated by qualitative case studies of cyber proxy relationships (e.g., Maurer, 2018a) with some recent quantitative analyses (e.g., Maness and Valeriano, 2016; Steiger *et al.*, 2018). The reason we propose a quantitative analysis is to contribute to a methodological gap in the literature and observe whether a statistically significant generalisable pattern emerges from the data. To test our claims, we propose a regression analysis testing for a positive correlation whereby an increase in the number of cyber attacks is associated with an increase in the number of MIDs globally, controlling for delegation of cyber attacks. Thus, we can compare outcomes between delegated and non-delegated attacks on the number of MIDs.

Independent and dependent variables

The dependent variable is the total number of MIDs that occur globally, expressed discretely. The total number of MIDs was chosen as the dependent variable because the effects of one state increasing its cyber capabilities go beyond a single dyad; one state increasing cyber capabilities in response to a cyber attack increases the risk of MIDs among all dyads. This data will be obtained from the Correlates of War Militarized Interstate Disputes dataset. The independent variable is the total number of cyber attacks globally, expressed discretely, using the Dyadic Cyber Incident and Dispute Dataset by Maness and Valeriano (2016). This dataset records all rival state-to-state cyber incidents between 2001-2011.

Controlling for delegation

States do not divulge information about their delegation of cyber attacks. Hence, a heuristic for the total number of delegated cyber attacks will be constructed as follows. We can observe cases where attribution of delegation of a cyber attack was successful. For example, “Guccifer 2.0”, a Romanian hacker who claimed credit for the hack of the Democratic National Committee in 2016, later discovered to have been a smokescreen for the involvement of the Russian government (Sanger *et al.*, 2018). We can also observe non-attributed cases whereby ambiguity about the origin of the cyber attack still exists. These cases cover the “uncertainty” aspect of the causal mechanism. Therefore, the sample of delegated cyber attacks will consist of successfully attributed cyber attacks using a cyber proxy in addition to non-attributed cyber attacks. The dataset will be built by monitoring news articles focusing on cybersecurity. After identifying relevant incidents, we will supplement the data using official government documents; reports from the IT-security community; reports from international organisations that analyse international security implications of emerging technologies; and national newspaper coverage, in order to gather as much information as possible to code incidents of cyber attack delegation.

Other control variables

The study should control for several other potentially relevant confounders. One factor that may influence the severity of a cyber attack and exacerbate the risk of an MID is type of damage. This will be expressed as a binary variable, coded as “1” if the cyber attack had kinetic effects and “0” otherwise. Other confounding factors include interaction type (e.g., espionage, an offensive attack); target type (e.g., a government base); and severity type (e.g., minimal effect, effect on infrastructure). These variables will be expressed categorically.

CONCLUSION

Proxy warfare as an instrument to achieving political aims is nothing new. These dynamics have sparked a budding literature on the strategic advantages and dynamics of forming cyber proxy relationships. In conventional proxy war, states often delegate because the proxy is able to carry out fighting at a lower cost than the state itself. The added-value of delegating cyber attacks to proxies lies in the state’s ability to evade attribution, given the difficulty of this endeavour in cyberspace. Thus, we hypothesise that an increase in uncertainty around the strength and actions of other states seems to increase the risk of MIDs, which is only exacerbated by the use of proxies, especially in cyberspace.

Notwithstanding, a correlation between the number of MIDs and delegation of proxy may not imply causation. There is always the possibility that incidents will be overlooked in our dataset due to the classified character of cyber attacks. Notwithstanding, we believe our dataset takes account of the genuine difficulty of attributing a cyber attack to an identifiable actor. The methodology furthers understanding of the effects of delegated cyber attacks on

MID risk, which may significantly improve the understanding of interstate dynamics as well as international conflict and stability.

Our causal mechanism describes the increased risk of MIDs among all dyads because we assume a decentralised network of states. Future research could model various network structures using simulations to see how they influence MID dynamics (Barabási, 2016). Research on how divergences in political systems affect cyber proxy relationships could provide further insight on MID risk.

BIBLIOGRAPHY

Barabási, A.-L. (2016) *Network Science*. Cambridge University Press.

BBC News (2014) ‘Kenya breaks “Chinese-run cyber crime network”’, 4 December. Available at: https://www.bbc.com/news/world-africa-30327412 (Accessed: 11 January 2021).

Berkowitz, J. M. (2018) ‘Delegating Terror: Principal–Agent Based Decision Making in State Sponsorship of Terrorism’, *International Interactions*, 44(4), pp. 709–748. doi: 10.1080/03050629.2017.1414811.

Borghard, E. D. and Lonergan, S. W. (2016) ‘Can States Calculate the Risks of Using Cyber Proxies?’, *Orbis*, 60(3), pp. 395–416. doi: 10.1016/j.orbis.2016.05.009.

Byman, D. and Kreps, S. E. (2010) ‘Agents of Destruction? Applying Principal-Agent Analysis to State-Sponsored Terrorism’, *International Studies Perspectives*, 11(1), pp. 1–18. doi: 10.1111/j.1528-3585.2009.00389.x.

Dunn Cavelty, M. (2013) ‘From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse’, *International Studies Review*, 15(1), pp. 105–122. doi: 10.1111/misr.12023.

Fearon, J. D. (1995) ‘Rationalist explanations for war’, *International Organization*, 49(3), pp. 379–414. doi: 10.1017/S0020818300033324.

Forsyth, J. W. and Pope, B. E. (2014) ‘Structural Causes and Cyber Effects: Why International Order is Inevitable in Cyberspace’, *Strategic Studies Quarterly*, 8(4), pp. 112–128.

Gartzke, E. and Lindsay, J. R. (2015) ‘Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace’, *Security Studies*, 24(2), pp. 316–348. doi: 10.1080/09636412.2015.1038188.

Gomez, M. A. (2019) ‘Past behavior and future judgements: seizing and freezing in response to cyber operations’, *Journal of Cybersecurity*, 5(tyz012). doi: 10.1093/cybsec/tyz012.

Gomez, M. A. (2020) ‘Overcoming uncertainty in cyberspace: strategic culture and cognitive schemas’, *Defence Studies*, 0(0), pp. 1–22. doi: 10.1080/14702436.2020.1851603.

Gross, M. L., Canetti, D. and Vashdi, D. R. (2017) ‘Cyberterrorism: its effects on psychological well-being, public confidence and political attitudes’, *Journal of Cybersecurity*, 3(1), pp. 49–58. doi: 10.1093/cybsec/tyw018.

Hansen, L. and Nissenbaum, H. (2009) ‘Digital Disaster, Cyber Security, and the Copenhagen School’, *International Studies Quarterly*, 53(4), pp. 1155–1175.

Hoffman, W. (2019) ‘Is Cyber Strategy Possible?’, *The Washington Quarterly*, 42(1), pp. 131–152. doi: 10.1080/0163660X.2019.1593665.

Jarvis, L., Macdonald, S. and Whiting, A. (2017) ‘Unpacking cyberterrorism discourse: Specificity, status, and scale in news media constructions of threat’, *European Journal of International Security*, 2(1), pp. 64–87. doi: http://dx.doi.org.libproxy.ucl.ac.uk/10.1017/eis.2016.14.

Jones, D. M., Bremer, S. A. and Singer, J. D. (1996) ‘Militarized Interstate Disputes, 1816–1992: Rationale, Coding Rules, and Empirical Patterns’, *Conflict Management and Peace Science*, 15(2), pp. 163–213. doi: 10.1177/073889429601500203.

Libicki, M. C. (2011) ‘Cyberwar as a Confidence Game’, *Strategic Studies Quarterly*, 5(1), pp. 132–147.

Liff, A. P. (2012) ‘Cyberwar: A New “Absolute Weapon”? The Proliferation of Cyberwarfare Capabilities and Interstate War’, *Journal of Strategic Studies*, 35(3), pp. 401–428. doi: 10.1080/01402390.2012.663252.

Maness, R. C. and Valeriano, B. (2016) ‘The Impact of Cyber Conflict on International Interactions’, *Armed Forces & Society*, 42(2), pp. 301–323. doi: 10.1177/0095327X15572997.

Maurer, T. (2018a) *Cyber Mercenaries: The State, Hackers, and Power*. Cambridge: Cambridge University Press. doi: 10.1017/9781316422724.

Maurer, T. (2018b) ‘Cyber Proxies and Their Implications for Liberal Democracies’, *The Washington Quarterly*, 41(2), pp. 171–188. doi: 10.1080/0163660X.2018.1485332.

Mumford, A. (2013) ‘Proxy Warfare and the Future of Conflict’, *The RUSI Journal*, 158(2), pp. 40–46. doi: 10.1080/03071847.2013.787733.

Powell, R. (2004) ‘Bargaining and Learning While Fighting’, *American Journal of Political Science*, 48(2), pp. 344–361. doi: https://doi.org/10.1111/j.0092-5853.2004.00074.x.

Reed, W. (2003) ‘Information, Power, and War’, *American Political Science Review*, 97(4), pp. 633–641. doi: 10.1017/S0003055403000923.

Reiter, D. (2003) ‘Exploring the Bargaining Model of War’, *Perspectives on Politics*, 1(1), pp. 27–43. doi: 10.1017/S1537592703000033.

Rid, T. (2012) ‘Cyber War Will Not Take Place’, *Journal of Strategic Studies*, 35(1), pp. 5–32. doi: 10.1080/01402390.2011.608939.

Salehyan, I. (2010) ‘The Delegation of War to Rebel Organizations’, *Journal of Conflict Resolution*, 54(3), pp. 493–515. doi: 10.1177/0022002709357890.

Salehyan, I., Siroky, D. and Wood, R. M. (2014) ‘External Rebel Sponsorship and Civilian Abuse: A Principal-Agent Analysis of Wartime Atrocities’, *International Organization*, 68(3), pp. 633–661.

Saltzman, I. (2013) ‘Cyber Posturing and the Offense-Defense Balance’, *Contemporary Security Policy*, 34(1), pp. 40–63. doi: 10.1080/13523260.2013.771031.

Sanger, D. E., Kirkpatrick, D. D. and Perlroth, N. (2017) ‘The World Once Laughed at North Korean Cyberpower. No More. (Published 2017)’, *The New York Times*, 15 October. Available at: https://www.nytimes.com/2017/10/15/world/asia/north-korea-hacking-cyber-sony.html (Accessed: 11 January 2021).

Sanger, D. E., Rutenberg, J. and Lipton, E. (2018) ‘Tracing Guccifer 2.0’s Many Tentacles in the 2016 Election (Published 2018)’, *The New York Times*, 16 July. Available at: https://www.nytimes.com/2018/07/15/us/politics/guccifer-russia-mueller.html (Accessed: 11 January 2021).

Slayton, R. (2016) ‘What Is the Cyber Offense-Defense Balance?: Conceptions, Causes, and Assessment’, *International Security*, 41(3), pp. 72–109.

Smeets, M. (2018) ‘The Strategic Promise of Offensive Cyber Operations’, *Strategic Studies Quarterly*, 12(3), pp. 90–113.

Steiger, S. *et al.* (2018) ‘Conceptualising conflicts in cyberspace’, *Journal of Cyber Policy*, 3(1), pp. 77–95. doi: 10.1080/23738871.2018.1453526.