National Strategies during Cyber Conflict: Secrecy versus Publicity

Dissertation Summary

Gil Baram

September 2020

In its 2020 Global Risks Report, the World Economic Forum ranked cyber attacks among the top ten risks in terms of likelihood and impact. Given the covert nature of cyber attacks, this concern is neither new nor surprising. Cyber warfare technologies enable countries to act covertly. Even if an attack has openly visible consequences – such as shutting the national power grid or telecom systems – the victim may claim it was the result of a technical malfunction. Likewise, the attacker may deny any connection.

Cyber attacks between states have been traditionally shrouded in governmental silence. due, in part, to fundamental technological and political challenges surrounding attribution. Indeed, inasmuch as any information about cyber attacks was ever made available, it mostly came from civilian sources, such as private cyber security companies (in the Stuxnet case), or journalists (as with Sandworm). Typically, governments remained silent on the topic.

Scholars and practitioners alike have, for the most part, consistently perceived cyber attacks as covert actions, obscured from public view because of the inherent characteristics of the associated technologies, and due to a (mis)perception that only technology professionals can understand them. According to this approach, the technological characteristics of cyber attacks allow concealing their origin, making attribution a key problem. Indeed, this paradigm has dominated most scholarship on the topic that emerged during the past decade.

In recent years, however, silence is giving way to public attribution, in which state agencies and governments both acknowledge and attribute cyber attacks. Despite the apparent advantages of maintaining secrecy on cyber attacks, states increasingly disclose involvement. This suggests that our understanding of strategic interactions between attacker and victim in this context is limited both theoretically and empirically.

Cyber attacks now seem to be emerging from the shadows into the public domain of international diplomacy and warfare, openly discussed by attackers and victims alike. I argue that countries – both victims and perpetrators – win political gains when they choose to reveal the attack. This choice of public strategy is not a binary political decision between revealing or concealing the attack. Rather, as I show, there is a variance in available strategies, ranging from complete silence to public attribution or credit claiming. Going public is not a zero-sum game; nor, as I demonstrate in this research, is it an arbitrary one.

Literature

Despite burgeoning research on cyberwarfare, scholarship has rarely questioned the premise for secrecy. Nor has research recognized the role of publicity in the aftermath of cyber attacks, and the advantages states stand to gain from publicly acknowledging them.

In the current state of the art, secrecy has been a defining feature of cyber space and operations within the cyber realm. Today it seems that offensive cyber operations have become a routine in countries' arsenal of covert activities. The covertness of cyber attacks can be expressed in two ways. First, the attack itself is covert. Its technological characteristics enable an attacker to carry out the operation in a clandestine manner, without revealing the existence of an attack, how it was conducted or what were its real objectives. Second, even in case the occurrence of an attack has been revealed, the attackers can keep their involvement secret.

The literature offers three mechanisms explaining why countries would choose covert actions. First are sunk costs, which refer to situations where states decide to take covert action because of non-recoverable resources. By choosing covert actions, leaders can employ a more "creative" way to address security threats. Second are counter-escalation risks. Using covert action can appear credible because of its impact on the risk of crisis escalation, since leaders using covert signaling tools can be free to engage in more aggressive behavior. This explanation is based mainly on the audience costs theory, which identifies a link between the type of action that the state takes and the costs the leader would have to bear as a result. The third mechanism is signaling resolve: under certain conditions, the use of covert operations allows states to convey the desired message to their rivals, thus circumventing the need to act in the public arena. When it comes to cyber attacks, these theories would lead us to expect that countries would remain silent and benefit from secrecy, a key characteristic of cyber warfare technology. However, data from recent years shows this is not always the case.

The secrecy component in cyber space often makes it difficult to identify the source of an attack and attribute it to a particular attacker, a problem commonly referred to in the literature as *the Attribution Problem*. Most research addressing the attribution problem discusses the meaning and significance of anonymity and the difficulties it poses to the attacked country when dealing with the consequences of the attack and deciding how it should respond. To date, the attribution problem is regarded as a fundamental challenge to governments, a twofold problem that is both technological and political. Technical means alone are insufficient to determine the motivation for an attack.

I posit that the attribution problem is not as critical as is commonly argued, and suggest moving forward from addressing it in mostly technical terms. Rather than asking who carried out the attack and how, we should be examining who will in fact be blamed for it and at what political consequence. In practice, countries routinely accuse each other of attacks without disclosing their technical attribution processes. On several occasions, for example, such as the Sony hack (2014) and the "WannaCry" attack (2017), the US publicly blamed North Korea.

This topic introduces political and geo-strategic facets heretofore understudied and which are crucial to understanding the true impact of attribution problems in shaping countries' strategic choices during cyber conflicts. The two key bodies of literature that examine cyber attacks in this context—Covert Actions and Cyber Warfare (the latter within International Relations)—separately predict silence on the part of governments in the aftermath of a cyberattack. Yet, they work in parallel and do not explore the nexus where these different disciplines actually intersect in reality, creating a theoretical gap: The first does not address the unique characteristics of cyber attacks and tends to exclude cyber operations from its analyses; the second largely accepts secrecy as an immutable feature of cyberspace rather than a strategy actors can opt into or forfeit. Why and when countries choose to exchange the secrecy widely presumed advantageous during or following a cyber attack with a different strategy has to yet to be fully examined in literature.

This is precisely the key focus of my dissertation, the main goal of which is understanding the phenomenon of countries choosing to forgo this advantage and disclose their cyber status in the international arena. Studying this phenomenon calls for new and updated **theoretical** explanations, as the change in attribution practices occurs within the theoretical gap identified at the juncture between Covert Actions and Cyber Warfare. Furthermore, this research is the first to deal with the questions of secrecy by systematic **empirical** examination, while marrying those separate, but deeply related, bodies of literature. This twofold approach is the theoretical anchor of my research.

The Theoretical Framework

The theoretical approach underpinning this study is a combination of theory and analytics, which come together to serve as an integrative framework. The theoretical component explains the the range of strategies. My analytical component identifies the variance in the strategies of both victim and perpetrator, ranging from complete silence to public statements at varying degrees of detail. This twin-pillared integrated analytical framework focuses on explaining the available strategies, with particular attention to the logic and reasons that led me to challenge the widely held belief that states are expected and predicted to maintain secrecy during cyber attacks.

Tradeoffs between the strategies are critical for the analysis of the strategic choice to abandon the advantages of secrecy in favor of a public strategy. As not all countries choose to either publicly reveal the attack or to hide it, the strategies of the victims vary between four possible approaches:

(1) Maintaining ambiguity – denying or downplaying any damage, thus reducing the chances that the attack would ever be revealed;

(2) Revealing damage – disclosing damage but denying it was caused by a deliberate hostile attack (claiming technical malfunctions, system "glitches" etc.);

(3) Admitting injury – publicly acknowledging that an attack took place, while refraining from identifying an attacker;

(4) "Pointing a finger" – publicly disclosing that an attack occurred *and* publicly putting the responsibility on a specific attacker;

Conversely, when victims choose the options of admitting to being attacked *and* choosing to accuse a specific rival as their attacker, the accused party has three response strategies:

(1) Ambiguity – remain silent and offer no public response whatsoever;

(2) Public denial – publicly deny any knowledge of or involvement in the attack;

(3) Credit claiming – taking public responsibility for the attack.

States, I argue, would willingly choose to relinquish secrecy and go public when the political and geostrategic circumstances warrant such behavior. And, I further put forth, this would happen much more frequently than commonly thought.

Choosing any of these options may have implications for the state, its leaders and its relations in the international arena. Therefore, it is important to examine in depth the calculations that may influence decision-makers to choose each strategy, the possible factors that led to the choice, and the various implications of the strategy ultimately chosen.

The **attacker** may choose a strategy of maintaining ambiguity for three reasons: First, when it wants to avoid the exposure of sources and its modes of action; Second, when it chooses to convey a message in a discreet and non-public manner; Third, when it wishes to allow the victim space to avoid having to retaliate.

Sometimes the attacker would prefer not to remain silent but to opt for a public denial strategy. There are three possible reasons for this: First, fear of retaliation. In cases where there are gaps in military capabilities between the countries involved, public denial may help reduce the likelihood of counter-action; Second, public denial may reduce the likelihood that the home audience (and even the international community) will express its dissatisfaction; Third, a desire to start an engagement and change the status quo. An example of this is China's repeated public denial of its attacks against the United States. This led the US to engage in a years-long process with China that ultimately resulted in the two countries signing a mutual agreement in 2015 to refrain from cyber attacks for intellectual property.

Contrary to the above two options, the attacker may choose the third strategy of public credit claiming. This may happen for two possible reasons: First, the desire to deter the victim; Second, to create a reputation of cyber-power by exposing capabilities. These two reasons are interconnected because creating a reputation for cyber power allows for deterrence. However, in light of the separate references in the literature to the topics of deterrence and reputation, this division was also applied here.

The **victim** may choose a strategy of maintaining ambiguity for three reasons: First, when wanting to avoid the obligation to respond and escalate the conflict; Second, when wanting to learn the opponent's course of action without providing information about the extent of the attack's success; Third, the victim may opt to claim malfunction when the attack has been exposed by a third party and cannot be concealed.

The victim may choose the strategy of exposing the attack when it wishes to avoid damage to its reputation that might be incurred in the event that the attack is exposed by the attacker or by a third party. The choice to expose the attack prevents the attacker from setting the media agenda and shaping the narrative of the attack in its favour. The victim may also choose to expose the attack without revealing the attacker for three reasons: First, when wanting to avoid the exposure of intelligence and capabilities; Second, to avoid glorifying the attacker's achievements; Third, to avoid the potential outcome of escalating the conflict.

On the other hand, the victim may choose to expose the attack and attribute it to a specific attacker when wishing to expose the attacker as violating international norms, and when seeking to use the threat posed by the attack to justify government investment in capacity development, or as a justification for future counter action.

My theoretical framework explains the range of strategies—from complete silence to public statements—of both attackers and victims. I then use an originally complied dataset on state attribution of cyber attacks in the past two decades to empirically examine this framework. The empirical analyses categorically refute the null hypothesis that states exclusively prefer secrecy and public silence on cyber attacks.

This research studies five hypotheses that examine the correlation between the political characteristics of the countries and targets involved, and the likelihood of them choosing a particular strategy: 1. Attackers with low levels of democracy will tend to publicly deny the attack; 2. Attackers with high levels of democracy will tend to reveal the attack and claim credit for it; 3. Victims with higher levels of democracy will tend to expose the attacks and even expose the attacker and attribute the attack; 4. When an attack has caused severe damage to the attacked state’s national security, victims will choose public strategies rather than maintaining ambiguity. 5. When the targets attacked are military or of other national security importance, both sides will tend to remain silent regardless of the countries’ level of democracy.

Research Design

In order to establish the empirical anchor of the research, I compiled a new dataset of all known state-related and state-sponsored cyber attacks between 1996 and mid-2019. This includes both officially attributed cases, as well as cases in which the attacker remained formally unnamed but which experts estimated were carried out by a nation state. An example of the latter was the 2018 attack against SingHealth, Singapore's largest health provider, described as the country's worst cyber attack.

First, I used the general framework of the *Dyadic Cyber Incident Dataset* (DCID) v 1.1. This dataset includes data on cyber attacks between countries for the years 2001-2014, and is based on the accepted coding of the *Correlates of War project* (COW). Using the COW framework, which serves as the main quantitative framework for international relations scholarship and research, will allow my postdoctoral research to extend the analysis to the wider geo-politics of cyber conflict. In order to improve our understanding of the political aspects of the attribution problem, I collected and analyzed the cyber conflict data so that every incident listed in the DCID 1.1 was examined and recoded based on the new variables and measurement scales I developed, such as who exposed the attack and the attacker, when, and more.

Second, to the DCID 1.1 framework I added data for the years 2005-2019, based on the Council on Foreign Relations *Cyber Operations Tracker* dataset. Published in 2017, this dataset includes a running list of cyber attacks from 2005 and is updated every three months, currently including data reaching mid-2019. Here too, I examined each incident separately and recoded it according to my new variables and scales.

Third, I added the relevant data for the years prior to 2001, based on the National Security Archive Cyber Vault Project at George Washington University. This project forms a public repository of primary documents obtained under the Freedom of Information Act and from other sources, and enables deep investigation of cyber attacks and cyber operation from the past decades.

In addition to the existing datasets and documentation cited above, I compiled additional data from mainstream media such as The New York Times and The Wall Street Journal, as well as professional cyber magazines and technical reports by cyber security companies, in order to update the dataset with incidents and threat actors revealed more recently, and recode these as well. Additional sources included threat intelligence reports, conference presentations, and social media. It is important to stress that all raw data collected for this dataset is open source. Once complied from the various databases and sources, all data was updated and recoded according to the variables I developed, thereby creating a new dataset with new capabilities. Upon completion of the data collection and coding, a logistic regression model (both binomial and multinomial distribution) was applied in order to examine the variables that influence the attackers and victims' choice of strategies.

In addition, I selected five case studies for an in-depth analysis. Chosen according to clear criteria, each case allows examining different aspects of the choice of countries to maintain secrecy or give it up. The cases are: The 2010 Stuxnet attack against Iran that was attributed to the United States and Israel; The Russian attack against the Democratic National Committee in 2016; The US attack against ISIS in 2016; the attack against Singapore’s SingHealth in 2017; The collision event of the US Navy destroyer John S. McCain with an oil tanker in the Malacca Strait in 2017.

Results

The research confirms that there is in fact an empiric departure from previous state silence on cyber attacks. Contrary to the prevailing assumptions in the literature that attribute many benefits to maintaining secrecy, this research shows that countries chose to give up secrecy much more often than expected.

The results of this study confirm the research hypotheses and, together with the results of the case studies, enable a better understanding of the reality in which countries operate during cyber attacks and the strategies they choose in the various cases. The findings indicate that there are clear features in the attackers’ strategies. While maintaining ambiguity is the attackers’ preferred strategy, in about one-fifth of cases, the attacker prefers to give up ambiguity and secrecy through taking credit or at least publicly denying the attack.

The victims’ strategies show greater variance than the attacker: While in nearly half of the cases, the victim chose ambiguity rather than address the attack publicly, in the second half of the cases they chose a public strategy. The publicity is expressed in two ways: Exposure of the attack without attributing it to a particular attacker, and exposure of the attack including public attribution. A partial disclosure of revealing the damage only accounts for less than 4% of the cases. The data demonstrates that the victim does in many cases choose to expose the attack, and that while maintaining ambiguity remains a very common strategy, it is not the only one available and that in many cases the attacker chooses to expose the attack and the attacker.

The statistical analysis and case studies confirmed the first and second hypotheses that when the attacking country has a low level of democracy it will tend to publicly deny the attack, and when the attacking country has a high level of democracy it will tend to expose the attack and take credit for it. In addition, the findings confirmed the third hypothesis that victims with democratic characteristics will tend to expose the attacks and even expose the attacker and attribute the attack. Also confirmed was the fourth hypothesis, that when the damage caused to the national security of the state as a result of the attack is severe, the victim will choose public strategies and will not tend to maintain ambiguity. The findings also partly confirmed the fifth hypothesis ($0.05<p< 0.1$) that when targets of military or other importance to national security were attacked – both sides will tend to refrain from disclosing the attack.

The research results are important for several reasons, not least of which because they demonstrate an empiric change on the ground in governments' cyber strategies in the international arena, one which scholars and policy makers currently lack the tools to study and address. This changing international arena requires new methodologies as well as practical instruments, and my research has laid the innovative groundwork for this new field.

In sum, state silence on cyber attacks is often replaced by public acknowledgement, which highlights the complexity of political attribution and informs key debates in political science, international relations and cybersecurity. What is more, as government officials increasingly engage in behavior of this sort, the implications of this research should be of interest to practitioners as well.

Contribution

In this study I identified a new variance in the strategies of countries that has not been identified in previous studies, and allows for a more accurate understanding of what is happening in reality and the considerations of countries in managing cyber conflict. The uniqueness of this study lies in providing insights into the connection between countries' political characteristics and their strategic behaviour during cyber conflict by identifying patterns of action of both the attacker and victim. This is one of the first studies to do so, providing tools for political scientists to create a better connection and understanding between the world of cyber attacks, which until recently was considered secret and inaccessible to many, and political and geo-political characteristics of countries.

Most of the literature on cyber warfare in International Relations is theoretical and qualitative, and the amount of quantitative studies is still relatively small. Using a combination of qualitative and quantitative research methods, this study offers a more comprehensive and broader understanding of the phenomenon and allows for more accurate theoretical and empirical conclusions, with higher generalizability and external validity.

This study makes a significant contribution to the possibility of using game theory models for decision making because it allows estimating the probabilities that countries involved in cyber conflict will take in each strategy. Thus, using the results of the study will allow decision makers to assess what the other side's strategy will be given its geo-political characteristics and the attack description.

The research findings carry two practical benefits: First, they identify the factors influencing states' decisions to give up ambiguity, concede being attacked, or publicly respond to accusations, and use statistical tools to mapping the range of strategic attribution options. Creating a theoretical infrastructure that allows making projections and identifying scenarios ex-ante will translate this academic approach into an innovative methodology that will serve future scholarship in this new field.

Second, they have practical utility for national decision makers and defense analysts. In particular, the findings will allow decision makers to: (a) decide on the optimal strategy to adopt under different circumstances during cyber conflicts; (b) predict their rivals' most likely strategy; and (c) consider the most efficient way to respond to it.