The Ethics of Cyber Warfare

Lina Dayem

University of Chicago

Introduction

As modern society advances technologically, information networks have become vulnerable to wrongdoing by malicious states and non-state actors. With recent strikes affecting critical infrastructures around the globe, the threats associated with cyber attacks no longer seem like science fiction. While the world has not yet faced catastrophic cyber assaults, our dependency on information networks exposes us to potentially devastating attacks. These technologies present attractive targets for cyber attackers aiming to undermine national interests or even to threaten state sovereignty. The international community is at a critical juncture: we now have the opportunity to determine what is morally permissible with regard to cyber warfare before we are ever faced with a worst-case scenario.

This essay draws upon Just War Theory to examine the military responses that are morally permissible in the face of a cyber attack. Indeed, certain cyber attacks originating from a state's government can be considered acts of war when analogous to conventional attacks either in means or in effect. These cases may justify a self-defensive response from victim states. Cyber responses are preferable to conventional responses in these cases, depending on the victim's technological capabilities.

However, the realities of cyber engagement have particular qualities, which, in contrast to other forms of conflict, render these more straightforward ethical norms less applicable. Firstly, the most dangerous cyber attacks are not physically immediate in the way of traditional weapons. Thus, ethical norms based primarily on the permissibility or impermissibility of physical violence are less straightforwardly applicable to cyber attacks without considering the grave, physically harmful potential of targeting immaterial code. Secondly, and more importantly, cyber attacks are often difficult to credibly attribute. The epistemological problem associated with an unattributed cyber attack leaves its victim at a seeming impasse: if the state cannot credibly identify its aggressor, how can it justify a counter-strike?

This paper takes a different, less traditional approach toward the difficulty of attribution, as well as toward the justified responses to identified non-state actors. I argue that according to the present legal and military norms, the epistemological bar for justified military retaliation is set at a level that may be appropriate for conventional attacks, but inappropriately high for cyber attacks. While very precise attribution to the source computer(s) may not be possible in many cases, the state from which the attack originated can more readily be identified. I contend that if a cyber attack can be reliably traced to the territory of a particular state, this state should be held at least partially responsible for the attack. Calling for the establishment and enforcement of codified norms of domestic and international cyber criminality, I argue that if a state becomes a frequent launchpad for cyber attacks, does not reasonably cooperate with victims to identify perpetrators, and fails to enforce criminal laws prosecuting such attacks, the state may ultimately be liable for these attacks. If diplomatic means prove ineffective, victim states would be justified in a reprisal. This punitive form of retaliation would only be permissible in a narrow range of cases and should only be limited to temporarily disabling the launchpad state's cyber testing capabilities.

Are Cyber Attacks Acts of War?

If cyber attacks can be categorized as acts of war, then a conventional attack may be a justifiable response. Certainly, the idea of using physical force against attacks on computer systems may seem counterintuitive. However, I argue that these intuitions stem from the fact that many cyber attacks thus far have not exhibited the physical characteristics of conventional attacks, making straightforward classifications a challenge. However, the term "cyber attack" denotes a vast array of potential operations, many of which may be analogous to recognized acts of war.

According to international legal norms, the first use of force is prohibited. This is apparent in Article 2.4 of the United Nations Charter, which prohibits "the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." The term "aggression" denotes the first use of force, which would justify a victim state's self-defensive war. The term is defined in Resolution 3314 of the United Nations General Assembly: "Aggression is the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State." According to this document, aggression includes, but is not limited to (see Article 4): invasions, bombardments, blockades, and armed attacks by one state against another (see Article 3).

While purposefully non-exhaustive, the language of the document conveys the notion that war is physical, transgressing real boundaries and causing tangible effects. These two documents were drafted in the mid-20th century, so it is unsurprising that they qualify military coercion in a manner consistent with contemporary warfare: by its instruments of "arms" and "force" (Schmitt, 2010, p. 154). By contrast, the physical coerciveness of a cyber attack stems from its consequences, not from its instruments. At the same time, Resolution 3314 does not exhaustively define "armed attack," leaving an interpretive space where cyber attacks could fit. For instance, considering computers and digital code as weapons leads to a broad definition of aggressive cyber attacks. On the other hand, when assuming a more strict interpretation of "force," then the documents do not prohibit non-physical economic or political coercion (Tallinn Manual, p. 46)—consequences of many of the cyber attacks we have witnessed to date. Ultimately, aggression broadly includes threats to and breaches of the peace (UN General Assembly Resolution 3314, Preamble), and the explicit purpose of the United Nations is to "maintain peace and security" (Charter of the United Nations, Article 1.1). Thus, it is reasonable to believe that at least some physically coercive cyber attacks could map onto its definition of aggression, even by a relatively conservative interpretation.

However, without an established convention for classifying cyber attacks as a form of aggression, philosophers as well as military ethicists have proposed three main standards for analyzing whether a cyber attack can be classified as an act of war. The first is a "means-based" metric. This standard classifies a cyber attack as an act of war if the attack produces the same type of physically-immediate destruction that an existing conventional weapon can, thus mirroring existing military means. The second standard is "target-based." By this metric, a cyber attack constitutes an act of war if it damages national critical infrastructure.

The final standard is "effects-based," which classifies a cyber attack as aggression if it produces a physically violent or overall destructive consequence to its victim. Therefore, the type of harm itself may not be completely analogous to that created by conventional weapons. This metric regards injurious effects as those that either create physical harm (like the means-based metric) or engender unacceptable physical or digital interferences to critical infrastructure (like the target-based metric). As with the target-based metric, what exactly constitutes "unacceptable interferences" is ambiguous and open to interpretation.

Note that the last two metrics could, by certain interpretations, consider some harm that is not physically threatening (such as interfering with financial services) to be aggression. However, just because a type of cyber attack is considered aggression does not give a state a *carte blanche* to start a war. Indeed, a war started in response to such an attack would not meet the *jus ad bellum* proportionality requirement (i.e. the threshold at which the harm done to a victim justifies war as a proportional response), since the potential loss of life and damage to property would be unacceptable to defend one's state against an economic downturn. The necessity requirement (i.e. the state's need to resolve a conflict through war) may not be met either. Indeed, a war would not be an effective or immediate way to reverse an economic downturn, and would be much more likely to exacerbate it.

Regardless of which of these metrics is adopted, the crucial point is that scholars and international policymakers (most notably NATO and the US Department of Defense) do recognize that cyber attacks can and should be considered as acts of war. By extension, the use of force in self-defense may be a permissible response by victims.

What is a Just Response to a Cyber Attack?

A just response to a cyber attack will vary based upon the type of attack, as well as the entity that perpetrated it. The conduct with regard to a state or a non-state actor will entail different procedures. Attacks may be attributed to states or non-state actors, or they may go unattributed. Cases attributed to states are the most straightforward. Most actions that constitute aggression would justifiably prompt a victim state to undertake a war of self-defense, provided that the *jus ad bellum* standards of proportionality and necessity are met. Responses against state-committed cyber attacks that do not constitute aggression may include "naming and shaming" or economic sanctions.

Cyber "aggression" attributed to a non-state actor cannot be considered an act of war because, according to international law, only states can declare war upon each other. Therefore, attacks of this character should be considered cyber criminality, and would require international cooperation between the victim and the "launchpad" state to pursue the attacker. Certainly, if an attack cannot be attributed at all, the victim state cannot react. However, as I will argue later in this essay, non-attributed attacks that can be reliably traced to a particular territory may justify certain types of force in response.

In what immediately follows, I will discuss permissible conduct with regard to attributed attacks. For the sake of argument, I will define aggression using the "effects-based" metric because it has been adopted publicly by the US and NATO, and consequently has real-world policy relevance. However, I acknowledge that if this metric of aggression is interchanged with any of the other standards, the permissible conduct in each case may change as well.

Illustration 1: State-Attributed Cyber Attack with a Violent Effect

In this case, state X has launched a cyber attack on state Y, targeting an automated weapon on a base in the territory of state Y, causing it to activate and fire at a false target within

the territory. On the surface, this case appears analogous to a UAV being flown over the border of state Y, or artillery shell being fired over the border into the territory of state Y (Strawser, 2010, p. 354). However, a key difference is that no enemy person nor weapon violated the territorial integrity of state X. Indeed, the hijacked weaponry itself originates in the attacked state—it originates in state Y rather than in state X—even though the computer initiating the attack may be in a remote location.

This nuance, while noteworthy from a tactical standpoint, does not create confusion when it comes to the internationally conventional "aggression." Consider the UN General Assembly's definition of aggression. It states that aggression entails the "use of armed force to deprive peoples of their right to self-determination, freedom and independence, or to disrupt territorial integrity" (UN General Assembly Resolution 3314). Specifically, we are told that "Bombardment by the armed forces of a State against the territory of another State or the use of any weapons by a State against the territory of another State" qualifies as aggression (UN General Assembly Resolution 3314, Article 3b). Paying close attention to the article's language, we note that the locus of the attack's origin is not specified. Therefore, there is no categorical difference if the attack is initiated in state Y or elsewhere; rather, the *effect* of that attack must be within the territory of state Y. Furthermore, the article uses broad language when referring to weapon types: it takes any weapon into account. Thus, the definition of aggression does not preclude a cyber attack, since in this case, it activated a conventional weapon, and a computer has been mobilized to a violent end.

Now that state X's attack qualifies as aggression, let us assess how state Y may proceed. According to the UN's definition, "A war of aggression is a crime against international peace. Aggression gives rise to international responsibility" (UN General Assembly Resolution 3314, Article 5.2). Taken together with the Charter of the United Nations, and assuming that a peaceable resolution cannot be forged, then it would be legally justified for state Y to respond to state X's attack through military means, since state X has begun an illegal war against state Y, and Y has the right to defend itself against this attack. Indeed, the document specifies that "nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations" (The Charter of the United Nations, "Chapter VII: Action with Respect to Threats to the Peace, Breaches of the Peace, and Acts of Aggression," Article 51). In short, a military response may legally be launched against a cyber attack with a violent effect.

Illustration 2: State-Attributed cyber attack with a plausible hostile threat

In this case, imagine that state X has launched a cyber attack on state Y, targeting state Y's government servers. The attack causes a shutdown of the system of military control and command of state Y, temporarily interfering with military communication. Since we are classifying aggression with an "effects-based" metric, it is evident that this attack constitutes aggression: an attack of this magnitude on a security system would constitute a serious attack on a state's critical infrastructure, thereby justifying a self-defensive response.

Some observers may object to the idea that a conventional attack would be justified to this form of aggression. However, I reject this position. Unlike case 1, this cyber attack produces no immediate physical harm. We may consider this attack a cyber form of the military tactic of interdiction. In conventional circumstances, interdiction is defined as "an action to divert, disrupt, delay, or destroy the enemy's military surface capability before it can be used effectively against friendly forces or to achieve enemy objectives" (Scott, 2016, p. vii). Applying this definition to the present case, state X executes cyber interdiction on state Y. Notably, the phrasing of the above definition does not pertain only to a first use of force. Indeed, it may imply that the state upon which the interdiction was carried out was already in a state of war or had prior hostile intent. Therefore, to avoid the causality dilemma of attributing state X's behavior to preemption, I assume for the sake of argument that state Y gave state X no reason to believe that it was planning an imminent attack, nor threatening violence.

It is plausible for the government of state Y to believe that state X has hostile, even bellicose, intent. According to the UN definition, "threats" on peace may be considered aggression. One such threat detailed in the document is a blockade of ports or coasts (UN General Assembly Resolution 3314, Article 3c). A blockade is similar in character to a cyber interdiction insofar as it does not have an immediate violent effect, but disrupts a state's capabilities. In fact, blockades are often categorized as interdiction (Scott, 2016). The categorical similarity between cyber interdiction and blockades may be enough to argue that state X's act is aggression, and allowing state Y to legally proceed as in case 1. Once again, a military response may be permissible against a cyber attack.

However, some may argue that cyber interdiction is not analogous to a blockade because of the very evident imminence that a conventional blockade implies. Without this direct link to the UN Charter, the question still stands: would it be plausible for state Y to believe that state X poses a hostile threat? To answer, I will invoke Walzer's logic of preemption. Walzer uses the purposefully vague term "sufficient threat" as the defining trigger of preemption (Walzer, 1977, p. 81). While Walzer does not give a comprehensive list of the types of attacks that would constitute sufficient threats, he does offer a set of reasoning allowing us to judge different situations. On his rule, "states may use military force in the face of threats of war, whenever the failure to do so would seriously risk their territorial integrity or political independence. Under such circumstances it can firmly be said that they have been forced to fight and that they are the victims of aggression" (Walzer, 1977, p.84). A response to a sufficient threat would be considered an act of self-defense.

The disabling of military control and communication would seem to pose a sufficient threat in Walzer's sense. This is because State Y could logically assume that state X's intention was to inhibit their ability to effectively mobilize against an unknown threat. Based upon Y's prior knowledge and relationship with X, this situation may breed a high level of fear in Y, leading them to anticipate any number of frightening scenarios. In this case, if Y has reason to believe, based upon contextually relevant factors, that X has dangerous hostile intent, then X poses a sufficient threat. Thus, a preemptive strike is permissible (that is, if state Y is able to

mobilize some sort of counter-strike, despite the attack). Since this attack is defined as self-defense against unjust aggression, then a military response is justified, by the logic of self-defense employed in case 1. What aim would a preemptive strike serve against X when the exact threat it poses is unknown? A physical or cyber attack targeting X's military-related critical infrastructure or X's command and control could undermine X's conventional and/or cyber capabilities. This strike could exacerbate escalations if mishandled. Therefore, it should only be undertaken if it is reasonable to believe that the attack could succeed. If successful, the strike may allow Y to take control of escalations, and thwart X's unknown, future attack.

Illustration 3: State-Attributed cyber attack without violent effect or hostile threat

In this case, imagine that state X has launched a cyber attack on state Y, targeting voting machines and altering election results. For the sake of argument, assume that neither the elected candidate nor opposing candidates planned to start a war, or commit atrocities such as genocide or enslavement. Considering the act in isolation, there is not enough information to determine state X's motivations in tampering with the results. Since no candidate claimed hostile intentions, we cannot say that state X was trying to thwart an election result that threatened one or multiple nations. Nor can we say state X wanted to ensure that a potentially violent candidate came to power. It is undeniable, however, that state X violates state Y's right to self-determination by tampering with the election results. A violation of self-determination alone may be a dubious reason to go to war.

Moreover, the language of "effects-based" aggression is no longer applicable here, for there is no physical harm to persons or property or critical infrastructure. While it is surely morally objectionable to interfere with a voting system, the system does not qualify as critical infrastructure. Even if voting systems qualified, an attack in response to election tampering would not meet the necessity or proportionality criteria. For, Y could simply annul the results of the election without engaging militarily with X. The potential loss of life or property damage resulting from a strike against X cannot be justified. Therefore, a military retaliation, whether conventional or cyber, would not be permissible. A better course of action would be for state Y to reclaim the self-determination it temporarily lost through a revote (preferably in a manner that is not susceptible to cyber interference). State Y may also be justified in employing a non-military type of punishment against X, such as economic sanctions.

Illustration 4: Attacks Attributed to a Non-State Actor

The nascent international precedent with respect to cyber criminality, as evidenced in the most relevant international treaty, the Budapest Convention on Cyber Crime, is that states should be expected to cooperate with each other in the maintenance of global cyber security (Council of Europe, "Convention on Cyber Crime," 2001). International justice of this sort requires an obligation of states to pursue non-state actors who commit cyber attacks from the state's territory. (Graham, 2017) Beyond this, "it confirms the duty of states to prevent their territories from being used by non-state actors to conduct these attacks against other states" (Graham, 2017,

p. 94). Such obligations would include fortifying their cyber defenses, criminalizing cyber attacks within their own domestic law, finding and monitoring belligerent hacking groups before attacks occur, cooperating with the victim state to locate perpetrators, or even extraditing a cyber criminal to the relevant victim state. These obligations are reasonable, because they help to maintain an ordered cyber terrain and ultimately minimize the use of physical force.

On my view, the reasonable level of cooperation should be determined on a case-by-case basis, since states may have different levels of wealth and technological capabilities, which may be due to structural factors beyond their control. It may be that weak states are willing to cooperate. They genuinely may place a strong effort into their cooperation, but still lack sufficient capabilities to prevent attacks or pursue assailants. It would be unfair to punish such states for negligence, especially if they are willing to have the victim state aid them in pursuit of assailant. In the same vein, I contend that states willing to cooperate, but unable to uphold these obligations due to lack of resources or technological capability, should be given aid to fortify their cyber systems against attack or increase monitoring capabilities. This way, upholding the terms of the obligations will not be based primarily on wealth and technological advancements. The overall effect of such aid will be increased global cyber security.

However, if a state is unwilling to perform these obligations, victim states may reasonably believe that the attack was state-sponsored or endorsed. Moreover, it is possible that the state knew about a threat posed by a non-state actor, but did not act within its capabilities to thwart the threat, making the state culpably negligent. In these cases, victims may then be justified in imposing some sort of punishment or sanction against the state. Depending on the lethality of the attack or frequency with which attacks originate from that state, victims may be justified in holding this safe-haven state responsible for the acts committed by non-state actors. The permissible resources in this case mirrors that of non-attributed attacks: a one-time reprisal with the intention of punishing that state. I will argue for this type of response in detail later in the essay.

Self-defense: Conventional Responses or Cyber Responses?

When it is appropriate to respond to cyber aggression, a state may employ a cyber or conventional form of retaliation. In this section, I assess these two means for their relevant moral differences. Bare in mind that any morally justified response, whether cyber or conventional, would be subject to the constraints of proportionality and necessity.

Cyber attacks have the potential to minimize harm in several ways. Firstly, engaging in a cyber response minimizes risk to a state's soldiers because they do not have to physically be present in hostile territory, where they may become subject to enemy attack or capture. Secondly, cyber attacks have the potential to precisely target a specific area of code within a certain system without causing unnecessary damage to persons or property. Therefore, in theory, cyber attacks could effectively disrupt enemy systems while eliminating collateral damage. Conversely, even precise conventional means, such as UAV strikes, create excess collateral damage. In a similar vein, the degree of engagement can be controlled more easily with a cyber

attack than with a conventional attack. For instance, the visceral shock and immediacy of bombing a military base may engender a more rapid and more violent escalation than temporarily shutting down military communication lines. Finally, many cyber attacks are meant to be temporary (due to a time-sensitive code), or can be reversed with repairs or patches. On the other hand, the effects of most conventional attacks cannot be reversed. And while buildings can be rebuilt and populations can regrow, the actual damage caused by conventional attacks is permanent.

Despite the fact that a cyber attack may be enacted by anyone, a highly complex cyber attack takes an advanced level of technological sophistication to perform. At the same time, for an attack to be successful, it will often involve both cyber and conventional forms of reconnaissance and espionage (Wheeler and Larsen, 2003). Due to these factors, the entities that can successfully carry out a highly sophisticated attack may be limited to wealthy governments with strong technological and intelligence capabilities. Therefore, if we determine that only cyber retaliation is permissible against cyber attacks, we may inadvertently create an asymmetrical moral environment. Strong governments could enact cyber attacks against weaker governments without fear that they will succumb to a symmetrical response.

Therefore, I suggest that cyber attacks should be preferred to conventional attacks because they can minimize the harm associated with retaliatory attacks. Indeed, if a cyber attack could be effective enough to achieve a certain outcome, the necessity requirement may bar a state from utilizing a conventional attack in its stead. However, it may be permissible to enact a conventional attack in response if the victim state does not have the technological capabilities to respond with a reasonably effective cyber attack.

The Problem of Attribution

Some cyber attacks are claimed by their perpetrators at the outset, and other unclaimed attacks can be attributed by their victims, although the attribution may not be immediate. However, given the relatively low threshold for being able to commit a cyber attack, and the ease with which an attack's origins can be purposefully obscured, attribution becomes difficult. While some argue that the problem of attribution is not unique to cyber attacks (Cook, 2010), cyber attacks are particularly susceptible to attribution problems in a way that conventional attacks are not.

While Internet Protocol (IP) addresses can be traced—i.e. the specific code assigned to each device on a network—doing so does not always provide credible leads. For example, an IP address can easily be faked using proxy servers. Certain attacks use malware to infect "civilian" computers, turning them into robots to enact remote commands. Thus, tracing an attack to its computer of origin does not provide information about the computer that triggered the attack. One attack of this category, the Denial of Service attack, triggers multiple, even thousands, of computers in diverse locations to launch attacks simultaneously, making pinpointing the originating computer even more difficult.

Notably, in countries, such as North Korea, where the computer systems are centralized or heavily regulated, then it is reasonable to assume that a cyber attack was launched by, or directly commissioned by, the government (Cook, 2010). On the other hand, an IP traced to a state with a prominent group of non-state actors could serve as a smokescreen for government-launched attacks (Cook, 2010).

In recent years, cyber security experts have developed more sophisticated methods to get closer to attributing attacks, such as linguistic analysis (Boebert, 2011), tracing the pattern of the malware infection (Sklerov, 2009), or analyzing the attack's targets and its level of sophistication. While these methods are often inconclusive, they may be helpful in allowing us to pursue the most just response to a non-attributed attack. This is why certain experts consider attribution as a sliding scale of confidence, rather than pursuing a standard of 100% certainty (Jones, 2017; Wheeler and Larsen, 2003).

At the present moment, completely and decisively resolving an attribution problem requires dispensing time and money in conventional forms of investigation or espionage (Dipert, 2010). But we can imagine situations where a state faced with violence must react quickly to a cyber attack, even with the epistemological barrier imposed by the problem of attribution.

Can a state respond to an unattributed attack?

The attribution problem complicates cases 1 and 2 for two reasons. Firstly, aggression is defined in Just War doctrine as a crime of states upon states (UN General Assembly Resolution 3314). Both justifications above hinged upon the concept of self-defense against aggression. These justifications do not apply in the case of a non-state actor. Therefore, it would be insufficient to argue Just War Theory's doctrine of self-defense to aggression alone to justify an attack against a non-attributed strike. Secondly, if a state is to launch a counter-attack against a non-attributed cyber attack targeting the territory where it originated, it has to accept the possibility that in some scenarios, it may be attacking a non-liable community. In other words, if a non-state actor is the true perpetrator, then the attacking state breaks the other's peace, effectively committing aggression.

Yet, determining that the victim state cannot act against a violent attack or a hostile threat, from a consequentialist stance, may unintentionally generate a precarious precedent for coercive engagement. For if a state has no recourse in the face of the problem of attribution, it leaves itself open to many future attacks. The only time that it can counter is against assailants who are too careless or too ill equipped to cover their tracks (Eberle, 2013). Thus, we could inadvertently create a precedent where states and non-state actors alike would be motivated to develop systems that disguise their identities, knowing that the epistemological doubt they have created will leave their victims with limited permissible recourse. And the better malicious entities become at hiding their identities, the more dangerous and lethal operations they could commit without fear of detection or retaliation.

Against this backdrop, I reason that a victim state may be justified in holding a launchpad state responsible for an attack that is emitted from its territory. The purpose for this is threefold.

First, states will have less motivation to perpetrate acts themselves, since they cannot hide behind the smokescreen of non-attribution. Second, states will have incentive to control these entities by monitoring and policing within their borders if they know that they may be attacked for the actions of non-state actors. Additionally, they will become less inclined to harbor or fund these groups. Finally, this principle encourages states to become increasingly concerned for the maintenance of not only their own national cyber security, but also global cyber security: a posture that is appropriate to the fact cyber systems are at their very essence a mark of globalization. Therefore, from a consequentialist standpoint, holding states liable for attacks emanating from their territories would theoretically serve as an incentive against becoming a launchpad or safe-haven state, and at the same time, serving as a disincentive to becoming a cyber attacker, and promoting an overall safer cyber terrain.

If states are held responsible for the attacks, are they liable to violent retaliation? As discussed earlier in this essay, states have a duty to prevent cyber attacks from emanating from their territory, either committed by the state itself or by a non-state actor. A cyber attack traced to the state's territory indicates a failure in one of those two duties. The *ex post facto* investigation should determine if the state could have been reasonably expected to prevent such an attack from occurring. If the state could have prevented the attack, it becomes liable for the harm caused by that attack. Again, this is because, from a consequentialist perspective, punishing a state for failing in this duty would incentivize them to prevent cyber attacks.

To assess a launchpad state's liability for an attack, victims must look to the launchpad's domestic policy toward cyber attacks (Graham, 2017). The victim state should consider the launchpad state's criminal law, its cyber security fortifications and monitoring platforms, its record of cooperation with victim states in the past, and its record of arrest and prosecution of known cyber criminals (Graham, 2017). Strict criminal laws and rigorous law enforcement would be deterrents for cyber attacks (Sklerov, 2009). A lack of criminal laws or law enforcement may indicate a state's passivity and indifference toward preventing cyber attacks. The international community should give due vigilance to the potential of ill-intentioned states scapegoating innocent individuals in order to give the false impression that they comply with the law enforcement requirement. These factors, taken together, will help the victim to determine whether an attack could have been prevented from occurring. The level of reasonable cooperation expected from each state would differ based on its resources and technological capabilities. This should be taken into account so as not to create a precedent that would unfairly punish well-intentioned states lacking in adequate resources due to structural factors beyond their control. If an attack originates from such a state, it may be morally responsible, although not liable to punishment.

One may question why states should be held liable for attacks that they potentially did not commit. The debate about holding states responsible for the actions of non-state actors is not a new one. Indeed, it is often debated with regard to states in which terrorists hide. Some critics argue that holding a state responsible for the actions of non-state actors unjustly shifts the blame to an innocent state. Critics argue that the victim states are actually initiating hostilities by

unjustly invading an innocent state, thereby transgressing its sovereignty and territorial integrity. To this objection, I respond that states that allow attacks to launch from their territories are acting as a safe-haven for terrorists, thereby acting immorally. By failing to stem attacks, they increase the potential of harm to innocent people. Moreover, if a state cannot effectively police its borders, it demonstrates that it is not entirely sovereign. Retaliatory violence against such states would be punitive in character—punishing the launchpad states for not upholding their sovereign duty. Therefore, these states would be liable to a reprisal, as it is a punitive use of violence.

Reprisals

The doctrine of reprisals is a military convention that, with some key modifications, would allow victims to hold launchpad states responsible for any cyber attacks originating in its territory. By definition, a reprisal is "a limited and deliberate violation of international law to punish another sovereign state that has already broken [these laws]" (Partsch, 2000, p. 380-383). The doctrine of reprisals came under harsh scrutiny after WWII, and justifiably so, because it entails the purposeful targeting innocent individuals. Until this point, the doctrine of reprisals' rather straightforward, "eye for an eye" mentality was thought to intuitively appeal to fairness (Christopher, 2004).

Reprisals are considered to be punitive for two reasons. First, reprisals are used to punish the state for an unjust attack that it has committed. More importantly, reprisals are used to punish a state for failing in its sovereign duty to prevent its territory from becoming a safe-haven for belligerent non-state actors. This punishment is intended to be a one-time action to reestablish an already broken peace (Christopher, 2004; Walzer, 1977). Thus, a reprisal is meant to prevent an escalation to war, rather than initiate a new one. The reprisal is committed against the state for not being able to uphold, or refusing to uphold, this sovereign obligations of policing within its territory. These attacks are coercive ones, used to incite the state to autonomously police its territory (Walzer, 1977).

Reprisals are illegal under international law for two main reasons. Firstly, reprisals, being forms of punishment, constitute a form of retributive justice executed directly by states rather than by an international tribunal, etc. Secondly, reprisals generally involve the massacre of civilians or prisoners of war, both of which are illegal and immoral. However, with a key modification, I argue that a reprisal may be a morally justified response to a cyber attack in certain cases. A state's reprisal can be morally justified if it minimizes, and preferably eliminates, harm to morally innocent people. Walzer assumes this line of reasoning, stating that reprisals should target property. Reprisers should make certain that any bystanders leave the scene well in advance of the attack. Walzer justifies this stipulation by noting that attacks on property challenge state sovereignty, without committing an "affront to humanity" by harming individuals who have not forfeited their right to life in any way (p. 219). Thus, even if individuals were killed in the initial attack, doing so in the reprisal would constitute murder (p. 217).

In the case of cyber warfare, if we are committed to the notion that reprisers must avoid all harm to civilians, then reprisals symmetrical to the original attack would be unjust in many cases. Certainly, targeting the critical infrastructure of a state would generally be impermissible because it creates the massive potential for civilian harm. For instance, a reprisal against a cyber attack of a hospital's power cannot target another hospital because it would place civilians in unnecessary harm. Likewise, launching an artillery missile at civilians in order to reprise a similar attack would also be unjust. Thus, it becomes clear that to remain morally justified, the reprisal to a cyber attack should be chosen very carefully. However, the reprisal could target property, such as less consequential types of computer systems. Alternatively, the victim could disable the cyber attack testing capabilities of the other state. The property destroyed in the reprisal should be proportionate to what was destroyed by the initial cyber attack (Walzer, 1977). It is reasonable to assume that, as long as the harm is proportionate, the reprisal may be cyber or conventional. At the same time, there are important factors to consider when deciding between a cyber attack or a conventional attack, as explained in a prior section. We can transpose the same reasoning to reprisals: a cyber reprisal should be preferred to a conventional reprisal unless the victim state lacks effective cyber capabilities.

Thresholds to justify reprisals

It would be unjustified to undertake the use of force without attempting diplomatic means in advance. Commensurate to this idea would be the installation of a threshold by which to determine if states uphold their duties to prevent cyber attacks that originate from within their territories. I propose two such thresholds: one regarding the number of attacks originating from the state and the other regarding the degree of severity of these attacks.

The first threshold, the number of unattributed attacks that originate from a certain state, should be adopted because it demonstrates how rigorously a state upholds its duty to prevent itself from becoming a safe-haven for cyber attackers, or from being a repeat cyber attacker itself. If several attacks originate from the same territory, this could point to a few explanations, none of which are positive for the state in question. The first is that the state is unable to police within its own borders, indicating that it is not completely sovereign over its territory. The second is that the state is unwilling to install the necessary measures to prevent non-state actors from committing cyber attacks, meaning that it is harboring these attackers. Finally, it could mean that the state's government is committing the attacks itself, but has advanced enough capabilities to hide its identity, a prospect that is both dangerous and disingenuous. As stated earlier, well-intentioned, cooperative states that are weak for economic or structural reasons beyond their control would not be liable to attack. They may accept aid to fortify their cyber defenses or increase law enforcement capabilities. The actual threshold should be determined by the international community. When the threshold is reached, a victim may be justified in the use of force against the other state.

The second threshold, the severity of the attack, is an important one because it emphasizes the proportionality justification. An attack may not be severe enough to justify the

use of force against it. Furthermore, if an attack is extremely severe or threatens the critical infrastructure in a tremendous way, then a victim state may be justified in retaliating as a direct response to this single attack, but only after allowing the originator state a certain amount of time to attempt to rectify the situation, search for the assailant, etc. However, it is true that some very extreme attacks, for example, false activation of a nuclear weapon, require a far more urgent response than others. It is reasonable to assume that attacks taking a more extreme nature would be far more likely to mirror the bellicose motivations and advanced capabilities of a government than a non-state actor. The idea of being able to punish a state for a single unattributed attack is rather tenuous, and lends itself very easily to abuse. I stipulate that doing so should remain illegal, so as not to create a dangerous, easily-abusable norms, although in extreme cases such conduct may be morally justified in retaliating after a single attack.

Conclusion

The international community is at a crucial moment: we now have the opportunity to determine what is morally permissible with regard to cyber warfare before we are ever faced with a worst-case scenario. Arguably, this is the best moment to decide the moral principles, which will govern our future conduct by influencing policy determinations. In this essay, I have explored the just responses to a cyber attack arriving at one self-defensive account and one punitive account. Attributed attacks constitute a first use of force, and justify self-defensive responses by victims. I first assessed just responses to attributed attacks. These self-defensive responses varied based upon the attack's character, as well as the assailant. I then posited that cyber responses were preferable to conventional responses in these cases, depending on the victim's capabilities. Finally, I tackled addressed.

The problem of attribution requires a very different category of response, however. Victim states may sometimes hold launchpad states responsible for unattributed attacks originating in their territory. When states were liable to attack, then the punitive measure of reprisal was a justified response. Notably, I draw upon and adapt a traditional norm to address a very contemporary problem in this growing field of coercive engagement.

References

Boebert, W. Earl. "A Survey of Challenges in Attribution." *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy* (2011): 41-52. Web.

"Charter of the United Nations, Chapter VII." *United Nations*. United Nations. Web. 06 Dec. 2016.

Christopher, Paul. *The Ethics of War and Peace: An Introduction to Legal and Moral Issues*. Upper Saddle River, NJ: Pearson/Prentice Hall, 2004. Print.

"Convention on Cyber Crime." Council of Europe, 23 Nov. 2001. Web. <europarl.europa.eu>.

Cook, James. "'Cyberation' and Just War Doctrine: A Response to Randall Dipert." *Journal of Military Ethics* 9.4 (2010): 411-23. Web.

Dipert, Randall R. "The Ethics of Cyberwarfare" *Journal of Military Ethics,* 9:4 (2010), 384-410

Eberle, Christopher J. "Just War And Cyberwar." *Journal of Military Ethics* 12.1 (2013): 54-67. Web.

Frantzeskou, Georgia, Stefanos Gritzalis, and Stephen MacDonell. "Source Code Authorship Analysis For Supporting The Cybercrime Investigation Process." *Proceedings of the First International Conference on E-Business and Telecommunication Networks* (2004): 85-92. *INSTICC Press*. Web.

Graham, David E. "Cyber Threats and the Law of War." *Journal of National Security Law & Policy*. 07 Feb. 2017. Web. 26 July 2017.

Schmitt, Michael N. "Cyber Operations in International Law: The Use of Force, Collective Security, Self-Defense, and Armed Conflicts." *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*. 151-78. Print.

Scott, Kevin D. "Joint Interdiction." *Defense Technical Information Center*. Department of Defense. Web. 06 Dec. 2016.

Sklerov, Matthew J. "Solving the Dilemma of State Responses to Cyberattacks: A Justification for the use of Active Defenses Against States Who Neglect Their Duty to Prevent." 201 Mil. L. Rev. 1 (2009).

Strawser, Bradley Jay. "Moral Predators: The Duty to Employ Uninhabited Aerial Vehicles." *Journal of Military Ethics* 9.4 (2010): 342-68. Web.

Tallinn Manual on the International Law Applicable to Cyber Warfare: Prepared by the International Group of Experts at the Invitation of the NATO Cooperative Cyber Defence Centre of Excellence. Cambridge: Cambridge UP, 2013. Print.

"United Nations General Assembly Resolution 3314 (XXIX)." *United Nations General Assembly Resolution 3314 (XXIX)*. University of Minnesota Human Rights Library, n.d. Web. 06 Dec. 2016.

Walzer, Michael. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic, 1977. Print.

Wheeler, David A., and Gregory N. Larsen. "Techniques for Cyber Attack Attribution." *Defense Technical Information Center* (2003): *United States Department of Defense*. Web.