Ethical Challenges in Cyber Warfare: A Modular Evaluation of Offensive Cyber Justification

Jacob Shaha, Rebecca Marigliano and Kathleen Carley
Societal Computing, Carnegie Mellon University, Pittsburgh, USA

jshaha2@andrew.cmu.edu
rmarigli@andrew.cmu.edu
kathleen.carley@cs.cmu.edu

Abstract: Competition and conflict in cyberspace at all levels of society have become persistent in the modern world. As individuals and organizations are obliged or incentivized to engage in such competition -- either defensively or offensively -- understanding the ethical implications of cyber operations is increasingly essential. Ensuring actions in cyberspace are ethically coherent with actions in other arenas protects persons and organizations from cognitive dissonance. It can impose normative forces to keep cyberspace compatible with civil society as it is presently understood. Rather than applying extant and monolithic ethical frameworks to cyber operations, this paper explores a modular approach to ethical framework construction. We examine how cyber action might be justified by multiple broad ethical paradigms, determining how different traditions might shape ethical justifications and, therefore, the permissibility and scope of cyber actions. The paper focuses on the ethical justification for offensive actions by examining different case studies and ethical framework constructions, highlighting how the foundational decisions that define a person's or organization's ethical framework subsequently determine the scope of action permitted to or required of that entity. Ultimately, the paper seeks to reconcile the new conflict domain of cyber with longstanding ethical reasoning about conflict in general and to highlight the specific deviations or reconsiderations this new frontier may require.

Keywords: Cyber warfare ethics, Normative analysis in cyberspace, Offensive cyber tactics, Ethical theory composition

1. Introduction

Competition and conflict in cyberspace between nations, commercial entities, and individuals have become persistent conditions in the modern world (Floridi and Taddeo, 2014). As individuals and organizations are obliged or incentivized to engage in such competition -- either defensively or offensively -- understanding the ethical implications of such cyber operations is increasingly essential. Ensuring actions in cyberspace are ethically coherent with actions in other arenas protects persons and organizations from cognitive dissonance and can impose normative forces to keep cyberspace compatible with civil society as it is presently understood.

Substantial philosophical thought has been directed at the ethics of cyber conflict in the decades since the advent of the internet. The risk entailed is not up for debate: cyber conflict has already produced significant loss of knowledge, wealth, and even human life (Miller, 2022; Center for Strategic & International Studies, n.d.). However, the proper legal and moral framework for evaluating these actions and weighing the propriety of potential responses remains a point of active discussion.

Unfortunately, neither aggressors nor defenders can wait for philosophical consensus to implement and utilize these systems. Toward that end, organizations wishing to engage in cyber conflict ethically must find some rubric to bridge the gap between their ersatz ethical behavior outside of cyberspace and their potential actions within cyberspace.

The array of competing organizations and diverse AI/ML applications is nearly infinite and too broad for the scope of any one paper. We propose an examination of extrema as a starting point for the discussion, as conditions that hold in the extreme may then be refined to accommodate nuance. As such, this discussion focuses on nation-states as the deciding parties and is limited to actions that may (or may not) constitute *warfare* instead of *crime.* We distinguish the two primarily by the acting agent rather than the target: herein, crime is a cyber offense undertaken by private parties for their benefit. By exclusion, then, actions undertaken by nation-states – whether directly or directed through other parties – are within the realm of possible acts of war.

Our purpose is to consider *how* states currently justify their offensive cyber actions. We will identify common properties or values between explanations in examining these justifications. These commonalities will, in turn, illuminate guiding principles already in force as ethical guidelines for cyber actions, *even between nations with ostensibly different normative ethical frameworks.* In other words, we seek to identify the most common justifications for cyber offense and explore how significantly different guiding philosophies can produce the

same outcome. Finally, we offer a few critical considerations for policymakers as the cyber era continues to unfold to assist them in making cyber policy that is more intentionally ethical while retaining competitive advantage.

Our motivation is two-fold. First, the Western legal tradition places state conflict and private conflict on the same spectrum, seeking to regulate both through legal systems built by consensus and accumulated historical tradition and interpretation, and grounded in moral and ethical reasoning. Thus, identifying where and how cyber conflict may depart from existing tradition and reasoning is an essential step to incorporating ethical cyber action into the larger societal framework of conflict management. Second, some parties view cyber action as much more acceptable than "traditional" means of achieving the same end. A toxic combination of a lack of regulation, widespread vulnerability, inherent anonymity, unlimited reach, and the latent psychological comfort of not physically confronting the victim paint cyber offense as a seductively appealing way to achieve goals that may be less than legitimate quickly. Certainly, China and Russia's open use of cyber attacks, compared to their restraint in using conventional force, demonstrate their conviction that cyber offense is comparably acceptable. (Sherr, 2017; Scobell, Burke et al, 2020)

We must also address Just War Theory (JWT), the predominant ethical and moral framework through which modern Western powers have justified, regulated, and conducted armed conflict. The scope of this paper is not to arbitrate how cyber action may or may not violate *jus ad bellum* or *jus in bello* tenets; this is an open question with significant extant literature. Instead, we will assume that states undertaking offensive cyber actions consider those actions justified, seeing themselves as "in the right." This is debatable from incident to incident, but making such an assumption sidesteps the JWT argument by supposing that the combatants have already ruled under that framework and thus allows us to compare the motives and reasoning by which they reconciled their offensive actions to JWT's requirements. We will not, however, examine the specifics of JWT requirements relative to cyber action; our focus is on the larger motivating, justifying mindset of national leaders and not the nuance of post hoc justification.

We will consider examples of cyber offense against frameworks from five rough "categories" of ethical thought. While they are by no means exhaustive, we feel these categories cover the majority of reasoning underlying statecraft:

- *Teleological egoism*, or *realism,* in which agents make decisions to optimize their own utility, however, they may define and measure it.
- *Teleological utilitarianism,* in which agents decide to optimize a global notion of utility, however, they may define and measure it.
- *Kantianism*, used here to represent Kant's assertion of persons as inherent ends and his prohibition against using any person as merely a means to an end, thereby prioritizing human dignity over all other concerns.
- *Virtue,* in which agents make decisions based on attributes they purport, or aspire, to embody.
- *Contractarianism,* in which agents make decisions in accordance with a set of rules that are, or would theoretically be, agreed upon by all, or a significant portion of, the agents governed by those rules.

2. Case Study: Espionage

We define cyber espionage as offensive cyber actions involving unauthorized access to or seizing control of a target system, which results in the exfiltration of information. In these scenarios, the target experiences no immediate loss of function at the system or organizational level; their only loss is information.

To illustrate, we examine two notable cases of cyber espionage. The first is the Chinese-sponsored data breach of the US Federal Office of Personnel Management (OPM) in 2015 (Segal, 2016; Committee on Oversight and Government Reform, 2016); the second is the US' NSA-led covert surveillance of European government and corporate entities and individuals, revealed in 2013-2014 through a series of leaks and investigations (Von der Mark, 2019; Reuters Berlin, 2015). As stated previously, we assume that both nations believed their actions to be ethically justified and did not deliberately commit acts they believed or knew to be wrong (as judged by their own moral/ethical standards).

China does not openly acknowledge its offensive cyber actions, nor provide a direct justification for such, and so we must intuit their thought process as best we can. In the OPM hack, China targeted a competing nation to obtain information that might provide an operational advantage against that nation. Critical components of justifying this action could include:

- A realist belief that all nation-states are in constant contention, and no action is genuinely off-limits within that arena.
- The Maoist tenets that capitalists would never turn toward communism of their own volition, and that capitalists inherently seek to suppress or destroy socialist orders that threaten their system (Lowe, 1966).
- The belief that obtaining this information would deter US aggression, preempting possible aggressive actions.
- Traditional Confucianist interpretations calling for benevolent rulers to enforce a peaceful order and for benevolent peoples to rise against evil kings (Yao, 2011).

Underlying all these justifications are a few shared thematic threads:

- It is moral to act for the benefit of the Chinese community, even if against other communities.
- The inherent goodness of the Chinese system underwrites small wrongs if they contribute toward the greater good of spreading and strengthening that system.

Framed in this way, an advocate for this offensive cyber act could appeal to several categorical ethical traditions. In a standard act utilitarian sense, China's action is not wrong because it advances Chinese socialism, which produces the greatest possible good for the world. Conversely, in a Kantian sense, China's actions are justifiable writ large, as the data theft did not instrumentalize any human beings and was directly opposed to a system that *does* instrumentalize human beings (by their definition). However, the genesis of the attack was likely a phishing scheme or some other deception; in that instance, the attackers could be accused of utilizing the duped party as a mere end to their own goals. While Kant himself would condemn this act of deception, many subsequent adherents of his work argue that, if worded or presented correctly, even a phishing email could be ethically acceptable in service of a greater goal (Carson, 2010).

Contractarian frameworks can justify this action by limiting "rational persons" to those who recognize the value and supremacy of communism over capitalism; indeed, a core tenet of socialist agendas is that those clinging to capitalism are either oppressed, ignorant, or malicious exploiters. As previously mentioned, traditional Chinese virtue ethics, represented by the Confucianist and neo-Confucianist bodies of thought, urge loyalty to "family" before all else and can be interpreted to justify resistance in service of order.

The protracted NSA surveillance of allied and partner nations throughout the early 21st century offers a useful counterpoint to the previous example. Once again, the US never offered a formal justification for their actions (though the President did proffer multiple apologies) (Reuters, 2021), so we are again left to guess at the US rationalization under the same assumptions. These may include:

- Recognition that democratic nations contain many political movements, some of which are not friendly with the US or its interests; should any of these movements come into power, the US must be forewarned of hostile intent and agenda.
- Assertion that valuable intelligence is often collected incidentally, through "overheard" or second-hand means, and so casting a vast dragnet is necessary to effective collection; and as a corollary, the claim that many of the "enemies" of the US and Western society already live and operate within that society, using terrorist cells as an example.
- The assurance that the US would never action intelligence against a friend, and that collection of such is an incidental technical side effect of more extensive, noble purposes.
- The assertion that targeted nations are likely to collect against the US as well and that such is a common practice even among allies.

We must acknowledge that, as U.S. citizens, our perspectives on both China's and the U.S.'s justifications are inherently biased. This bias may lead us to attribute a greater unity of purpose or idealistic motivation to China than is warranted while possibly viewing U.S. actions with undue cynicism. The U.S. justifications, as presented, reflect a somewhat contradictory stance, alternating between two distinct views:

1. Geopolitics is fickle, and there are no permanent friends or enemies; all targets are valid as contingencies.
2. The US is inherently benevolent and can be trusted, even with sensitive access and information.

Essentially, "You can trust the U.S., even if the U.S. doesn't fully trust you."

This contradictory position presents a more significant ethical challenge as we seek an underlying explanatory framework. In some cases, because the US is positing the same paternalistic position as the Chinese did for communism, similar reasoning – however specious – can be used. In a utilitarian sense, the US would argue that the US-led national order is the best arrangement for the world, so any means required to uphold and defend it are ethically valid. Regarding human dignity, the same Kantian argument about carefully constructed deception and justified opposition to an unethical system that might have justified Chinese offenses can be applied to the US' actions.

However, purported Western ideals make other ethical frameworks more problematic. An Enlightenment-era contractarian view, for instance, cannot casually discard the rationality of Communists. Indeed, the US system allows for the existence of, and possible governance by, a US Communist party, should the electorate empower them; this allowance means that a contractarian must consider Communists as rational agents. In a framework where ethical actions are defined as only those to which all sensible parties would agree/consent, the presence of Communists at the bargaining table now rules out cyber offense against a nation-state based solely on ideology or possible futures (though *probable* futures may still be valid). A contractarian justification for US cyber offense must forego a unanimous consent requirement in favor of a partial framework. Because the US represents only 5% of the global population, any such framework will ultimately heavily weigh the US "vote" at the bargaining table.

Western virtue ethics may also pose a problem. Taking the Aristotelian tradition as an example, offensive cyber actions of this nature would appear to sacrifice loyalty and trustworthiness in favor of, at best, some patriotic ideal. Under the most widely understood version of virtue, it would be difficult to claim that lying to one's purported friends, even for one's own security, is ethically virtuous. Classifying "friend" as merely a label and not an inherently elevated relationship does not assuage the problem; it only makes the perpetrator seem less virtuous and more self-centered.

The cross-section of these examples, therefore, indicates that nations wishing to employ offensive cyber for espionage, regardless of their political/economic system, are likely reasoning:

- Under a teleological framework, they estimate that their system represents a more significant potential and/or realized suitable for all concerned;
- Under a Kantian framework, where they consider any deception to be acceptably conducted in service of a system that protects human dignity en masse; or,
- Under a cynical, realist framework, where ethics are inapplicable.

3. Case Study 2: Sabotage

We define sabotage as an escalation beyond espionage involving unauthorized use or access to a target system that directly destroys or leads to an immediate loss of capability or functionality. Here, "capability" and "functionality" are intentionally broad, encompassing the accessed system's specific abilities and the target organization's broader operational capacities. The crucial aspect of sabotage is that unauthorized access directly causes such loss.

Further, sabotage need not be closely linked in time to the resulting loss. We make no clear distinction between preparatory actions, such as establishing access (e.g., installing a covert backdoor), and executing an attack using that established access (e.g., corrupting data through a previously embedded backdoor). As the former is often a necessary precursor of the latter, when an attack does occur, inevitably all preparatory actions fall under the same ethical scope as the culminating act. A more interesting question is the ethical status of preparatory actions taken, but never leveraged; while no loss was inflicted, the perpetrator nevertheless maliciously penetrated the system. Taxonomically we may place such an action near enough to espionage to apply the previous section's logic, foregoing deeper consideration in this paper.

Sabotage generally falls into two categories, which we describe as "reversible" and "irreversible." These labels distinguish between acts of sabotage that are relatively easy to remedy—where the target must only expend minimal time, energy, or resources to restore lost functionality—versus those causing significant, lasting damage that requires substantial investments to repair or replace. Consider, for example, a radar installation as a non-cyber example. If a saboteur sneaks into the installation and flips circuit breakers, thereby cutting power to the radar, they have inflicted a loss of capability. However, that capability will be restored as soon as the owning entity diagnoses the problem and flips the circuit breaker back to "on," a series of actions that are (a) relatively simple and (b) completely restore the previous capability. By contrast, a saboteur placing

explosives along the antenna array and detonating them would inflict far more severe damage. In this instance, no trivial action can quickly or easily restore the installation's functionality. The owners must rebuild or replace the installation in part or whole.

Cyber sabotage often hovers near this boundary, since the investment required in cyber—in time, treasure, or talent—can vary significantly between nations. As such, we will *not* treat these two categories with any ethical distinction. For our purposes, we will consider the ethics of any cyber action that deprives a target of a capability the target finds valuable. As case studies, we offer the Stuxnet virus that targeted Iran in 2010 (De Falco, 2012) and the Russian cyberattacks against Estonia in 200 (Unlisted, 2007).

Though no nation has officially claimed responsibility for creating and deploying Stuxnet, conventional wisdom largely attributes its development to the United States. Assuming the U.S. considered this action ethically justified, we explore the likely rationale behind creating and deploying such a cyber weapon. Stuxnet destroyed Iranian nuclear production tools, depriving them of a capability into which Iran had invested significant time, energy, and cultural import; this was unquestionably an act of aggression. To justify what might have been a "first strike" in a thankfully avoided war, the US arguments likely included:

- A view of Iran as a hostile, unpredictable, and belligerent power, with the follow-on reasoning that Iran was likely to utilize nuclear weapons against the US and/or its allies if and when Iran came to possess these weapons.
- A belief that Iran would be unresponsive to additional economic and diplomatic efforts to slow or cease the pursuit of nuclear arms.
- A belief that the attack would go undetected or, if detected, would be unattributable.
- A belief that the attack would not result in the loss of human life, nor affect systems beyond the targeted Iranian centrifuges (though this later proved untrue) (Anderson, 2012).

JWT-based justification was certainly considered here, given the aggressive nature of this cyber action. Specifically, the primary US themes of justification seem to be:

- Violent or destructive actions limited to valid targets that inflict minimal collateral damage will not cause war and are, therefore, acceptable as part of the non-war competition.
- Preemptive strikes are acceptable as self-defense against hostile powers that would like to attack, given means and opportunity.
- Unattributed or undeclared violence is acceptable outside of war.

While point 1 aligns with JWT and point 2 is generally seen as nominally compatible with JWT, point 3 contradicts the JWT requirement of declared authority as part of *jus ad bellum*. This indicates that the U.S. justification for cyber offense originates in some ethical framework that is more permissive than the underpinnings of JWT. In some cases, the arguments that allowed espionage also permit this act of sabotage; the utilitarian case is the most obvious, so long as the assessed value of the US and the US-led international order exceeds the damage or potential damage suffered by Iran. Furthermore, by carefully designing a mechanism of sabotage that is destructive but expected to be non-lethal, advocates of this act can appeal to a Kantian framework in that they hold life sacrosanct and (outside of the previously discussed deception involved in launching the attack) have not trivialized or instrumentalized any persons, ensuring that the attack, if not completely "right," is at least definitively not "wrong."

As before, the consensus contractarian approach falters: no rational group would permit actions that might irreparably and unattributably damage capabilities in which they have made significant investments, which forces a contractual approach to exclude targets as rational participants. This position becomes even more difficult to sustain as the range of potential targets expands. Here a virtue approach is not quite as damning: since Iran has declared itself an enemy of the US, one might argue that what this action lacks in valor (as a covert action), it makes up for in wisdom and restraint. Virtue ethics do not necessarily require pacifism of adherents, and insofar as the US can be seen as a virtuous contributor in its struggle against Iran, Stuxnet – with its emphasis on non-lethal action and precise impact – might be the most virtuous example of state-on-state violence imaginable.

We contrast this case with the Russian attacks on Estonia. Unlike Stuxnet, Russia's sabotage, primarily a denial-of-service attack, falls mainly into the reversible category; most of the damage was promptly corrected, and the impact of the attack was measured more in confusion and lost time than in material loss. A crucial difference is the declared enmity between the two states or the lack thereof. While the US and Iran

acknowledged their differences, Estonia had no open quarrel with Russia, and Russia did not openly cite Estonia as an opponent. The attack was precipitated by Estonia's decision to relocate a Soviet war monument and cemetery against Russia's wishes.

Understanding Russia's justification for such bold and widespread cyber sabotage must account not only for the initiation of hostile action but for the deliberate souring of a relationship that might have otherwise been improved. Toward that end, we can infer Russia's justification to include:

- A moral duty to defend Russian and Soviet history and culture.
- A desire to influence other nations without resorting to physical violence.
- Confidence that NATO would not intervene or escalate to military action.

These justifications reveal significant moral themes for Russia:

- Russia has a leadership role over Slavic people and its region of the world.
- War is wrong, but conflict is acceptable; so long as Russia does not initiate open war, its actions are short of wrong.

Unfortunately, Russia's ethical calculus seems one-dimensional. Much of their justification appears to hinge on their perceived right or claim to the former Soviet republics and, indeed, to all territory they believe to be "traditionally" within their ethnic sphere of influence. Russian ethicists espousing a utilitarian view can claim evil in the service of the greater good based on a valuation of the Russian way over alternatives, as has been leveraged in previous examples. But a Kantian approach fails flatly. In attacking public services, Russia sought to influence the minds of a few leaders and decision-makers by influencing the suffering of the broader populace, thereby instrumentalizing thousands of Estonians. Even a generous interpretation of Kant's work cannot endorse statecraft through indiscriminate deprivation.

A contractarian view centered on Russian interests might be possible but would require excluding all non-Russian actors as rational agents; that would represent an ethnic supremacy position matching that of the Nazis, and so is unlikely to stand open scrutiny, even within Russia. However, a virtue-based framework may justify this attack for the same reason seen in the previous example. By avoiding bloodshed while achieving gains for the state, this action may espouse a virtuous form of patriotic combat, sparing lives while obtaining decisive victory. (That the "victory" achieved was far from decisive is inconsequential in the a priori justification; the potential for such a victory may be enough to convince a decision-maker of the rightness of this action.)

These examples, wherein states do actual damage to one another's capabilities – reversing or eliminating the other's investments, either temporarily or permanently – would seem to have a few possible justifying ethical frameworks. States undertaking cyber sabotage are likely making decisions based on the following:

- A teleological framework, where they estimate that their system represents tremendous potential and/or realized good for all concerned;
- A virtue-based framework, wherein aggression on behalf of the state is laudable when done with appropriate restraint; or
- A cynical, realist framework where ethics are inapplicable.

4. Case Study 3: Combat and Warfare

Having addressed espionage and sabotage cases, we now turn to acts of warfare. While it may seem straightforward to classify warfare as overt —since espionage and sabotage are covert— this is not necessarily the case. Many actions in declared warfare are still masked in order to prevent effective counteraction. In this context, *warfare* will mean a cyber offense conducted as part of a more considerable, *overtly declared* combat effort. This distinguishes it from the "open conflict" ideologies referenced previously. Although attribution challenges remain, the existence of avowed and ongoing hostilities would lead the target nation to attribute such attacks to the declared belligerent. This attribution can be problematic, as it may invite opportunistic actions by third parties—a longstanding tactic in warfare, now adapted to cyberspace.

We must also consider the case of a cyber "first strike" that *leads* to war. Our examinations of espionage and sabotage supposed that the perpetrators considered those acts *below* the threshold of initiating war. However, countless wars began over actions that the aggressor did not consider war-worthy, who saw the resulting conflict as a disproportionate reaction by the aggrieved. In the case of cyber conflict, such misunderstanding is no less likely to happen, and the murky boundaries between cyberspace and the natural

world may amplify the ambiguity in both parties' view of the offensive action. Although the prospect of a cyber first strike is intriguing, we do not specifically deviate to address it here for two reasons: first, there is no clear historical precedent wherein a cyber action led to open warfare (rather than *preceding* open and anticipated warfare); and second, there is no reason to believe the ethical justifications for such an act would deviate from those already examined, since the intent of the action (if not the result) was to avoid open war.

Our case study focuses on the Russian disruption of Ukrainian internet services in March 2022 during Russia's invasion of Ukraine (Coker, 2022). Other cases might be applicable; US jamming of cellular infrastructure in Iraq and Afghanistan were considered counter-cases. However, upon examination, we found the reasoning in both cases to fall along the same lines. While the Russian and US "way of war" may be significantly divergent at the tactical, operational, and even strategic levels, the two nations' conceptions of the *nature* of war are mainly parallel. Both sides recognize Clausewitzian war doctrines and construct their forces and campaigns around such historically proven tenets, albeit with different emphases.

Russia's large-scale coordinated cyber-attacks against Ukrainian ISP Ukrtelecom in March 2022 resulted in widespread internet outages across the nation, creating significant confusion within an already-panicked nation staving off an invasion. Interestingly, the justification for such an action is much narrower than in previous cases because Russia had already justified the broader umbrella of armed hostilities. This cyber-attack became one of many weapons used against an enemy in an already- "just" wider conflict, as opposed to an undeclared aggression with commensurate geopolitical risks. Even in open war, however, commanders must justify their selection of tactics and weapons against the objectives and circumstances of their force. To that end, Russia's specific and nested justification of these cyber-attacks likely followed regular operational lines of reasoning (Baxter, 1986), including:

- Mass (or, in Soviet doctrine, Concentration) – reasoning that a simultaneous attack by mobile forces *and* cyber effects would significantly impact Ukrainian forces.
- Economy of force (or the Soviet Capability) – reasoning that the cyber strike could have a widespread demoralizing and desynchronizing effect while exposing Russian forces to virtually no risk and requiring virtually no resources.
- Surprise (or the Soviet Initiative) – recognizing that a strike against national communications infrastructure would likely surprise Ukraine's high-level leadership and citizenry, leaving them scrambling to respond to those attacks and any follow-on events.
- Unity of Command (or Soviet Coordination) – recognizing that Ukraine's dependence on civilian infrastructure would make deprivation of the same debilitating to its military command and control, significantly reducing its ability to resist.

Because these rationales are along military operational lines, the underlying principles mirror longstanding military thought:

- Cyber offense is another category of military effects, like artillery or air-dropped leaflets, and can be woven into military operations accordingly.
- The Ukrainian military and government are operational targets. Those entities utilize the national communications infrastructure, which is a valid military target despite any civilian use and reliance on it.

Couching these principles within larger normative frameworks without slipping into a full-blown examination of JWT is difficult. We are intentionally limited in our consideration here. We can say that the attack's intentionally non-lethal and largely reversible nature appeals to Kantian and utilitarian lines of thought. We also argue that it appeals to a contractarian framework: supposing that the participants agree that warfare is sometimes unavoidable, we can imagine all of them endorsing a non-lethal, reversible means of attack over other possibilities. And for a virtue-based consideration, we see that, as before, what cyber offense lacks in heroic valor – risk to life and limb for a higher cause – it makes up for in prudence, efficiency, and cleverness.

Any modern warfighting nation in a state of open hostility will likely consider cyber offense justified under the broader umbrella of an ostensibly Just War. Concerns over dual-use targets and civilian suffering are no different in the cyber case than they are in the kinetic targeting of power plants and roads. Fears about the weapon exceeding its intended target are swept into the same conversation around incendiary weapons.

However, cyber weapons diverge in one crucial aspect: their persistence. Historically, tools like Stuxnet have demonstrated that cyber weapons can affect systems outside of their intended targets, and those effects have

lingered well after the desired impact was (or was not) achieved. In this way, some cyber offensive techniques may have less in common with artillery shells than they do with biological weapons or persistent chemical agents. Crucially, both means of warfare are international taboos; open utilization of such weapons, or proof of their covert use, might constitute casus belli for parties outside the original conflict, whose sole interest disempowers a regime willing to take such risks.

We do not argue that the possible existence of such tools justifies classifying cyber tools as "cyber bioweapons" and insisting that all parties forbear from using them. We do, however, recognize that a cyber attacker may at some point unleash a weapon – knowingly or unknowingly – that induces more effect or damage than the international community is prepared to abet. Because cyber operations lack the deterrent structure of nuclear arms, such incidents may become increasingly likely as cyber conflicts escalate. And, as previously mentioned, the guilty party will almost surely accuse the broader community of overreaction when they are vilified for a "mere" cyber-attack. We will not speculate about the nature of such an attack and argue that, as previously mentioned, the a priori justification for such an act would (unfortunately) likely not diverge from the lines of reasoning already explored.

5. Conclusion

Based on our considerations, it seems that nations willing to use cyber for espionage, sabotage, and/or open warfare must root their ethical reasoning in one of the following permissive frameworks, which were common to all three categories:

- A "paternal" utilitarian approach, wherein the nation must assert the clear superiority of its system over others, thereby justifying its actions as a net gain for humanity.
- A standard act egoistic approach is the cynical realist view that each nation must do what is best for itself before all other considerations.

This conclusion is a somber one. The overarching ethics of cyber-attacks would ultimately seem to be colored with delusional narcissism and dour fatalism. This may not be surprising considering the common ethical rationale of warfare in general. We might find loftier or more satisfying ethical underpinnings, extolling the right of self-defense, the duty to oppose tyranny, or the virtue of valor. Indeed, our examination of espionage and sabotage found more admirable possible frameworks. But the only reasoning systems that can justify cyber offense across all levels of conflict are, unfortunately, those that can justify *any* offense at all levels, and viewing such ideas in that light, it is difficult to see them as anything other than narcissism and cynicism. Thus, to justify cyber aggression, a nation must either shift its rationale to the needs of the moment or fully embrace one of a handful of stark overarching views.

Our conclusion is *not* that cyber offense is inherently unethical; instead, it is that cyber offense is *no more* ethical than *any other form of aggression*. Strategists and policymakers should not comfort themselves in resorting to cyber offense in pursuing their national interests, believing that moving the "battlefield" to a bloodless digital realm is more noble or justifiable. The reality is that, at the state level, the rightness or wrongness of a virtual attack is likely not substantively different than that of a physical attack. A state will fight to defend its survival and will imperil other states' survival in doing so. If a cyber-attack did not inflict such peril, it would not be effective as a tactic. That it does inflict peril means that targets will suffer and retaliate; the "virtual" nature of that peril is ultimately immaterial.

Cyber offense will be an indelible part of interstate rivalry so long as computer networks are an indelible part of interstate commerce and communication. Let us not fool ourselves into believing that the ill-defined "borders" of nations in cyberspace are somehow less sacrosanct than the equally arbitrary lines we have drawn on maps. A state launching a cyber-attack against another must justify it, in part, by admitting that, were it equally desirable and more convenient, they would likely launch a physical attack instead.

Acknowledgements

The research for this paper was supported in part by the Threat Assessment Techniques for Digital Data, Office of Naval Research under grant (N000142412414), the Army under grant (W911NF20D0002) through the AI2C center, and by the Center for Informed Democracy and Social-cybersecurity (IDeaS) and the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon University. The views and conclusions are those of the authors. They should not be interpreted as representing the official policies, either expressed or implied, of the the Office of Naval Research, the US. Army, or the US Government.

References

Anderson, N. (2012, June 1). "Confirmed: US and Israel created Stuxnet, lost control of it." Ars Technica. https://arstechnica.com/tech-policy/2012/06/confirmed-us-israel-created-stuxnet-lost-control-of-it/

Applegate, S. (2012). *The Principle of Maneuver in Cyber Operations.* 4th International Conference on Cyber Conflict. NATO CCD COE Publications.

Barnes, J. (2019, February 26). "Cyber Command Operation Took Down Russian Troll Farm for Midterm Elections." *New York Times.* https://www.nytimes.com/2019/02/26/us/politics/us-cyber-command-russia.html

Baxter, W. P. (1986). *The Soviet Way of Warfare.* Brassey's Defence Publishers, United Kingdom.

Carson, T. L. (2010). "Kant and the Absolute Prohibition against Lying." In *Lying and Deception: Theory and Practice.* Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199577415.003.0004

Center for Strategic & International Studies. (n.d.). *Significant Cyber Incidents.* Retrieved 18 October 2024, from https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents

Coker, J. (2022, March 29). "Ukraine Suffers Significant Internet Disruption Following Cyber-Attack." *InfoSecurity Magazine.* https://www.infosecurity-magazine.com/news/ukraine-internet-disruption-cyber/

Committee on Oversight and Government Reform, U.S. House of Representatives, 114th Congress. (2016, September 7). *The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation* (Majority Staff Report). https://oversight.house.gov/wp-content/uploads/2016/09/The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf

De Falco, M. (2012). *Stuxnet Facts Report: A Technical and Strategic Analysis.* NATO Cooperative Cyber Defence Centre of Excellence. https://ccdcoe.org/uploads/2018/10/Falco2012_StuxnetFactsReport.pdf

Floridi, L., & Taddeo, M. (Eds.). (2014). *The Ethics of Information Warfare.* Springer International Publishing.

Lowe, D. M. (1966). *The Function of "China" in Marx, Lenin, and Mao.* University of California Press, California.

Miller, M. (2022, December 28). "The mounting death toll of hospital cyberattacks." *Politico.* https://www.politico.com/news/2022/12/28/cyberattacks-u-s-hospitals-00075638

Reuters. (2021, May 31). "US Spied on Merkel and Other Europeans through Danish Cables – Broadcaster DR." *Reuters.* https://www.reuters.com/world/europe/us-security-agency-spied-merkel-other-top-european-officials-through-danish-2021-05-30/

Reuters Berlin. (2015, July 8). "NSA Tapped German Chancellery for Decades, WikiLeaks Claims." *The Guardian.* https://www.theguardian.com/us-news/2015/jul/08/nsa-tapped-german-chancellery-decades-wikileaks-claims-merkel

Scobell, A., Burke, E., et al. (2020) *China's Grand Strategy.* Ch 2. Rand Corporation.

Segal, A. (2016, February 16). "Why China Hacks the World." *Christian Science Monitor.* https://www.csmonitor.com/World/Asia-Pacific/2016/0131/Why-China-hacks-the-world

Sherr, J. (2017). "The Militarization of Russian Policy." *Transatlantic Academy,* Paper Series 10, 5-7. www.transatlanticacademy.org

Unlisted. (2007, April 28). "Tallinn Tense After Deadly Riots." *BBC News.* http://news.bbc.co.uk/2/hi/europe/6602171.stm

Von der Mark, F. (2019, September 17). "Snowden: Germany Typifies Surveillance Cooperation." *DW.* https://www.dw.com/en/edward-snowden-germany-a-primary-example-of-nsa-surveillance-cooperation/a-50452863

Yao, F. (2011). "War and Confucianism." *Asian Philosophy,* 21(2), 213–226. https://doi.org/10.1080/09552367.2011.563996