Cyber Operations and the Threshold for Cyber Warfare: Ethical and Anticipated Ethical Issues

Richard Wilson[1,2] and Noah Donnelly[2]
[1]Department of Philosophy, Towson University, Baltimore, Maryland, USA
[2]Computer Science and Information Sciences, Towson University, Baltimore, Maryland, USA

wilson@towson.edu
ndonnel1@students.towson.edu

Abstract: Determining whether a cyber operation meets the threshold for being designated cyber warfare involves ethical, technical, and strategic criteria. These are primarily derived from international law frameworks such as the Tallinn Manual 2.0 and principles of the UN Charter. What are the primary criteria for cyber operations to be called acts of Cyber Warfare Threshold? (1). Scale and Effects of cyber operations. The cyber operation must cause effects comparable to a kinetic attack, such as: Physical destruction (e.g., damaging power plants, pipelines). Loss of life or serious injury. Severe disruption of essential services (power, water, healthcare). Example: Stuxnet physically damaged Iranian centrifuges, meeting this criterion. (2). The Intent and Purpose of the Cyber Operation. The cyber-attack must be strategically motivated and aimed at: Weakening an adversary's military or economic capability, Coercing or punish a state. Achieving political or military objectives. Example: Ukraine power grid attacks were linked to geopolitical conflict. (3) Cyber operation must be attributed to a State. The cyber operation must be attributable to a state or conducted under its direction/control. State sponsorship or direct involvement elevates an attack from the level of being a cybercrime to the level of acts of cyber warfare. Example: WannaCry and NotPetya were attributed to North Korea and Russia respectively. (4) There must be severity and consequences related to cyber operations. The impact of cyber operation must be significant enough to trigger international legal obligations, such as a Breach of sovereignty. The possible classification of the cyber operation as a "use of force" under Article 2(4) of the UN Charter. Example: Triton/Trisis targeted safety systems in petrochemical plants, risking catastrophic damage. (5) The Target Type of cyber operation. Attacks by the cyber operation on critical infrastructure or military systems are more likely to meet the threshold. Civilian systems may also qualify if disruption is widespread and severe. (6) The Context and Scale of the cyber operation. The cumulative effects of cyber operations mean that: repeated or coordinated attacks may collectively meet the threshold even if individual cyber operations do not. This analysis will identify the ethical and anticipated ethical issues with the identification of Cyber Operations and the ethical criteria for these operations to reach the threshold for being classified as acts of Cyber Warfare. Our methodology will employ an interdisciplinary method that draws upon distinctions taken from Bratman's BDI Model of rational agency, computer science, and conceptual ethical analysis with reference to specific case studies.

Keywords: Cyber warfare, Tallinn manual 2.0, Article 2(4), Stuxnet, Attribution, Kinetic equivalence, Critical infrastructure, Gray zone conflict

1. Introduction

In classic Westphalian systems of international relations, what acts of war were often historically ambiguous. War was defined by physical acts with immediate and catastrophic consequences. Such actions could be armies crossing sovereign borders, a missile striking a city, or naval blockades. The physical violation of sovereign territory was the primary mode of starting a conflict. However, with the advent of the digital domain, what has now been introduced is a gray zone of conflict (Clement, 2018). The gray zone is a spectrum of hostility that exists below the threshold for armed conflict, but far above peace. This is where Cyber Operations lay, where states now regularly perform digital espionage, sabotage, subversion and coercion that cause quantifiable damage to their opponent without ever meeting the legal definitions of being called a war. The defining question facing the international community in the digital age is defining the exact moment at which a line of code is no longer a tool of intelligence, but an act of war (Kello, 2013). This paper argues for a fundamental paradigm shift in how international law evaluates cyber conflict, aiming to resolve current ambiguity. The current "Effects-Based" model waits for a cyber-attack to cause physical destruction equivalent to a kinetic weapon. We contend this approach is dangerously reactive and insufficient for the digital age. Instead, we propose a "Norms-Based" threshold. Under this framework, the act of war is defined not by the volume of damage caused, but by the nature of the target itself (e.g., hospitals, nuclear plants), regardless of whether the attack succeeds or fails.

The "Threshold Problem" is not just about semantics or academic debates, defining it is imperative to guaranteeing global stability. The definition of this threshold for cyber war would decide the rules of engagement for major powers on the global stage. If a state is able to define cyber-attacks on its power grid as an act of war, which in turn leads to claims the right to self-defense, this could lead to kinetic retaliation with conventional weapons under Article 51 of the UN Charter. This violates defense mechanisms of the escalation ladder where an act of digital espionage can spark a nuclear conflict. On the other hand, if states set a threshold

for cyber warfare that is too high, like demanding a physical act, adversaries would be allowed to erode a nation's sovereignty, steal intellectual property, and cause economic harm to a targeted nation with immunity. The criteria needed to establish this threshold are Scale, Effects, Intent, Attribution, Severity, Target Type, and Context. These are all derived from the *Tallinn Manual 2.0* on the *International Law Applicable to Cyber Operations* (Schmitt, 2017). This academic study represents a comprehensive attempt to apply existing international law to the cyber domain but fails to provide the application to clear up any ambiguity or interpretations.

This ambiguity creates a problematic ethical landscape where the rules for defining cyber warfare are decided by the aggressors. A logic bomb placed in a petrochemical plant, like Triton malware (Sobczak, 2019), is meant to cause destruction of infrastructure and loss of life, but if does not detonate due to an error in the code, international law struggles to classify what kind of act has occurred. Is it an act of war or an act that attempts sabotage? Does the effect of thousands of small hacks against the financial sector equate to a single large act that would be similar to a missile strike on a bank? This analysis will investigate these questions through the lenses of Just War Theory and Anticipatory Ethics. By examining case studies, including physical destruction caused by Stuxnet, the economic chaos of NotPetya which was the targeted disruption of the Ukrainian power grid, and the lethal intent of Triton. We argue that the current ambiguity is not an anomaly, but a feature of modern conflict. It is a feature which is exploited by state actors to project power while avoiding accountability. We conclude that ethical stability necessitates a transition from "Effects Based" definitions, which are reactionary in nature, to "Norms Based" which would prohibit targeting civilian infrastructure regardless of outcome (Farrell & Glaser, 2017).

2. Technical Issues

To determine a threshold for Cyber Warfare is complicated by the unique technical architecture of the internet, it resists the precise, binary categorizations of kinetic attacks. Unlike physical weapons, which are predictable in yield and trajectory, cyber weapons are often unpredictable in strength and attribution for cyber-attacks is difficult.

The main technical metric used by international lawyers to analyze cyber warfare is Kinetic Equivalence. Within the framework of the Tallinn Manual, a cyber operation qualifies as a "Use of Force" if its scale and effects are equivalent to physical operations while also rising to a level in which use of force is justifiable. Technically, this would require measuring the impacts of digital code on physical objects, people or computer systems. One example is the Stuxnet worm causing physical destruction of Iranian nuclear centrifuges by altering the way that the Iranian nuclear centrifuges operated in order to destroy them by manipulating the SCADA (Supervisory Control and Data Acquisition) systems monitoring the centrifuges. The bridge of cyber to physical damage is the clearest technical mark of warfare because it results in broken machinery. Many cyber operations do not break physical things; they merely alter or destroy data. The technical challenge is in the quantifying of equivalence between cyber and physical acts. If a ransomware attack such as WannaCry shut down the hospitals of a nation, causing no structural damage to buildings, but causing patient deaths due to delayed care and the inability to access medical records, is the alteration of code now capable of being classified as a weapon and does this rise to the level of an act of war? The difficulty lies in the attribution of causality. In a complicated and interconnected system, proving that a specific data packet caused a specific physical outcome is a challenge that continues to evade certainty (Rid, 2012).

A prerequisite for a declaration of war is identifying the enemy, but in the cyber domain attribution is a forensic challenge shrouded in obfuscation by the use of False Flags. In the physical world, a missile has a clear return address if its trajectory is tracked. Within the cyber domain, state actors can route their attacks through proxy servers that are in neutral countries, using purchased criminal malware to mask identity, or purposely insert snippets of code in foreign nations with the intention of framing other countries and masking their own identities. For example, the malware entitled Olympic Destroyer used against the Pyeongchang Winter Olympics contained code that looked like it was written by North Korean hackers (Lazarus Group), but in fact was produced by Russia's GRU which was discovered through forensic analysis (Greenberg, 2018). Technically, establishing the responsible party with a degree of certainty that would justify a kinetic military strike would take months to years of signals intelligence analysis. The "Time to Attribution" lag renders immediate "Right to Self Defense" inoperable in real-time. A nation cannot launch a retaliatory strike if eight months are required to determine who fired the 1st cyber shot.

In addition, state actors often employ tactics such as small-scale intrusions, that remain below the threshold of war to circumvent thresholds, but that collectively equate to a strategic assault. Detecting these activities

requires advanced Correlation Analysis over a wide array of disparate datasets. Scanning a port server in Ohio, phishing attacks sent to a Texas contractor, and a server outage in New York may appear unrelated, but in fact may not be. To pinpoint the link between these attacks would take a centralized cyber defense agency, recognizing that these events represent a coordinated campaign by a single Advanced Persistent Threat (APT). The inability of dispersed civilian networks to aggregate and correlate this data in a timely manner, prevents victim states from knowing they are at war until the campaign has caused maximum damage, and the perpetrators have disappeared. This tactic exploits the time lag of bureaucratic decision making, which offers little defense against the light speed efficiency of fiber optic networks (Rid & Buchanan, 2015).

3. Ethical Issues

Norms are being degraded because of the vacuum caused by the lack of clear definition for a threshold for cyber-attacks escalating to the level of cyber warfare. The application of Just War Theory to cyberspace to reveals inconsistencies between traditional military ethics and the realities of the cyberspace in which modern cyberwarfare is conducted.

The first ethical conflict arises from Sovereignty and Article 2(4) of the UN Charter. Article 2(4) prohibits the "threat or use of force against the territorial integrity or political independence of any state." The problem is that cyber operations often violate sovereignty without the use of physical force. The ethical issue lies in the notion of Westphalian Sovereignty being eroded. Russian interference in the 2016 U.S. election did not destroy voting machines, it violated the political independence of the United States in a significant way. If the threshold for war is based strictly on physical destruction or Kinetic Equivalence, it legitimizes non-destructive operations that become existential threats to a nation's survival rather than kinetic warfare and border disputes. This essentially provides a defense for psychological and political warfare, allowing states to destabilize rivals without any fears of kinetic military repercussions (Lucas, 2017).

The second ethical issue is Retaliation as defined in Article 51 of the UN Charter. Article 51 permits self defense in case of an armed attack. If a cyber operation crosses the threshold for this, the victim state is permitted ethically and legally to respond with a kinetic attack (Waxman, 2013). However, the attribution problem creates an ethical risk, Mistaken Retaliation. If state A hacks state B, but routes attacks through an innocent state C, and state B then attacks state C, innocents suffer due to errors in forensic analysis. Employing utilitarian ethics, the escalation from cyber conflict to nuclear or conventional attacks would suggest the threshold for attribution must be absolute. Demanding 100% certainty in attribution for cyber-attacks emboldens aggressors because they can now hide within the margin of doubt. This creates an ethical dilemma between defending the citizens of one's nation, and the duty to avoid starting unjust wars due to faulty intelligence.

Finally, the Principle of Distinction would need belligerents to distinguish between military and civilian targets. In cyberspace, this is basically impossible due to the fact that the military uses civilian internet for its operations. The backbone of the internet are fiber optic cables, routers, and DNS servers which are Dual-Use Infrastructure. Attacks intended to degrade an enemy militaries command-and-control network will in turn deny civilian internet access and affect banking and emergency services. Almost any strategic cyber warfare operation carries high risks of violating Jus in Bello (conduct in war). Accepting collateral damage in cyberspace has become normalized to a degree that is incredibly dangerous. States like Russia, as seen in the NotPetya attack, treated the global economy as a "free fire" zone. The ethical failure here lies in the refusal of states to separate military and civilian networks, which holds the civilian population hostage to military operations.

4. Case Studies

Boundaries for the Cyber Warfare Threshold have not been drawn by treaties, they have been drawn by precedents that occur in conflicts. These cases show how the threshold is tested, perverted, and defined by state actors which has led to arbitrary definitions.

4.1 Stuxnet

The Stuxnet worm which was deployed in 2010 and is attributed to a joint U.S.-Israeli operation, shows a paradigmatic example of a cyber operation that met the threshold of an armed attack. Stuxnet was designed with one purpose: the physical destruction of Iran's nuclear enrichment capabilities at Natanz (Zetter, 2014). The malware targeted Siemens Step7 software used to control Programmable Logic Controllers. It modified code sent to the frequency converter drives, which caused centrifuges to spin at supersonic speeds until they broke from mechanical stress. Simultaneously, the malware performed an attack on the monitoring systems by feeding

prerecorded "normal" data to the control room so the operators would not be aware of the destruction of the centrifuges on the plant floor.

A Threshold Analysis confirms that Stuxnet met the criteria of Scale and Effects for cyberwarfare by causing verifiable physical destruction. The "Target Type" criteria were met by attacking military infrastructure. While the Stuxnet attack met the threshold of an armed attack, the muted response of the international community represented a tacit acceptance of cyber sabotage, especially in regard to nuclear proliferation. Stuxnet presents a Utilitarian outcomes-based argument: the outcome of the attack delayed nuclear proliferation without causing human loss of life, framing it as a sanitary alternative to a kinetic airstrike. A Deontological duty-based analysis would argue that the criteria of "Scale and Effects" for cyber warfare was met by the undeniable physical destruction of the centrifuges, opening a "Pandora's Box," by validating the use of malware for physical destruction (Zetter, 2014).

4.2 Ukraine Power Grid Attacks

In 2015 and 2016 the Sandworm team, a group of Russian hackers compromised the SCADA systems of Ukrainian power distribution companies, leaving hundreds of thousands of people without electricity during the peak of winter (Greenberg, 2019). The attacks started with spear-phishing emails which contained "BlackEnergy 3" malware. The malware granted access to corporate networks from which the attackers pivoted to SCADA networks, acquired VPN credentials, and remotely accessed the control panels. The operators of power companies watched helplessly as their cursors moved across their screens, opening breakers and cutting the power to substations, all in a coordinated strike. Without a means to prevent the breakers from being switched back on, the attackers had to launch a "KillDisk" wiper component that erased the boot records of the operator stations, which would prevent any attempts to restore power remotely. On top of wiping the operator stations, they flooded the customer support center with Telephony Denial of Service (TDoS) attacks to prevent customers from reporting outages.

Threshold Analysis revealed that operation Sandworm was testing the "Severity and Consequences" criteria of cyber warfare. By causing no physical destruction or direct loss of life, NATO did not invoke Article 5 and operation Sandworm hovered firmly within the gray zone. By stopping short of permanent kinetic damage, the aggressor avoided triggering a war while causing adverse effects for the targeted population. The attack violated the ethical norm which protects civilian infrastructure, in this case power stations during peacetime, effectively treating the Ukrainian populus as a test subject for cyber weaponry (Greenberg, 2019).

4.3 NotPetya

In 2017, the Russian military intelligence (GRU) unleashed the NotPetya malware, disguising it as a legitimate tax software update for the Ukrainian accounting program. Unlike standard ransomware that encrypts a company's data for profit, NotPetya was designed to wipe data, an action that is purely destructive in nature. The malware utilized "EternalBlue" which is an exploit stolen from the U.S. National Security Agency (NSA), designed to spread rapidly through the Server Message Block (SMB) protocol of the Ukrainian accounting program. The original target was Ukraine, but the autonomous nature of the malware caused it to spread globally, infecting corporations like Maersk, FedEx, and Merck. The attacks costed an estimated $10 billion in damages, crippling global shipping and logistics for weeks (Greenberg, 2019).

Threshold Analysis of NotPetya obscures conceptual distinctions related to "Target Type" and "Intent." The intended target was Ukraine, but it became a global weapon of mass destruction which was due to its wormlike nature. The scale of the financial damage caused by NotPetya was equivalent to a kinetic war, yet every state ended by posturing, western nations publicly condemned the attack but stopped short of any meaningful retaliation. NotPetya is an example of the lack of accountability for intentionally and Indiscriminately deploying Cyber Weapons. The "spillover" effect violated the sovereignty of numerous neutral nations who were not party to the Russia-Ukraine conflict. The international community failed to classify this as an "Armed Attack" which indicates that the current threshold for a cyber-attack rising to the level of cyber warfare is too high, prioritizing the status quo over the economic health of the deeply connected global system.

4.4 Triton/Trisis

The Triton malware was a sophisticated cyber weapon built to target Safety Instrumented Systems discovered in a Saudi petrochemical plant in 2017. The unique danger of this malware lays in the targeting of the Schneider Electric Triconex Safety Instrumented System (SIS). The SIS is the last line of defense, intended to shut down a plant in case of dangerous pressure or temperature spikes to prevent an explosion within the plant. The malware

was intended to reprogram the SIS controllers creating unsafe conditions, effectively cutting the safety net protecting the plant (Johnson et al., 2017). The attack failed, but only because of a coding error in the malware which caused the safety controllers to enter into the failsafe state, shutting down the plant and alerting plant operators of the malware attack.

Threshold Analysis shows that the Triton malware showed an undeniable "Intent to Cause Loss of Life." Targeting failsafe systems is pointless if you are conducting espionage or intelligence work, its only purpose is to cause death and destruction. If not for an error, Triton could have caused a massive explosion and a release of toxic hydrogen sulfide gas which would kill workers and civilians in the surrounding area. This clearly crossed an ethical "Red Line" in civilized conflict. The coding error does not prevent the perpetrators from being blamed for the attack because their intent remains. This is a clear attempt at mass homicide via digital code, and represents a possible war crime (Sobczak, 2019).

4.5 SolarWinds

In late 2020, the "SolarWinds" supply chain compromise showed how fragile the threshold between espionage and an armed attack. SolarWinds was attributed to the Russian SVR, a foreign intelligence service, and was an operation that relied on stealthy infiltration. The attackers injected a backdoor known as "Sunburst" into the process of building software for the SolarWinds Orion network monitoring platform. The "Sunburst" code was digitally signed by a legitimate certificate which allowed it to bypass the standard trust verification in over 18,000 government and corporate networks. Once the attackers had established the backdoor, they used "Teardrop" which is a malware that moves laterally and exfiltrates data that is oftentimes sensitive. The unique challenge presented by SolarWinds is one of "Scale and Effects." The scale was massive, it compromised the Pentagon, DHS, and Treasury, the effect was not destructive in nature, no data was wiped, no systems were bricked. The Tallinn Manual 2.0 classifies this as espionage, which is not prohibited under international law (Schmitt, 2021). However, the ethical line is blurred due to the operation implanting malicious software into critical infrastructure, the physical equivalent of this would be preparing a battlefield. This blurs the line by questioning whether the act of war is merely the detonation of a bomb, or if planting explosives counts as an act of war. The SolarWinds case demonstrates the failure of current threshold analyses to take "Strategic Latency" into account, which is the capability to destroy being establish, but not immediately used (Microsoft, 2020).

5. Anticipatory Ethics

Failing to prevent these escalations in cyber attacks is a failure in foresight. Applying "Moral Responsibility for Computing Artifacts" rules by Miller et al. (2011) to design cyber weapons and targeting of critical infrastructure shows significant ethical lapses.

5.1 Rule 1: The Foreseeability of Effect

Rule 1 declares that designers are morally responsible for the foreseeable effects of their artifacts. This applies directly to state agencies developing "Zero-Day" exploits. While designing "EternalBlue" the NSA foresaw that if such a powerful tool were to be leaked, it could allow for bad actors to cause massive amounts of damage. When the Shadow Brokers leaked it, and when it was subsequently being used in the WannaCry and NotPetya attacks, the "foreseeable effect" later materialized in the form of billions of dollars in damages and risks to patient wellbeing in hospitals. Through the lens of Anticipatory Ethics, the idea is that as technology develops ethical problems will also develop, so stockpiling "Zero-Day" vulnerabilities for cyber-attacks is unethical at its core, because the state of the technology at the present time is that the weapon cannot be controlled. A nation state has an ethical duty to disclose any vulnerabilities to patch civilian infrastructure, this is a process called Vulnerability Equities Process, as opposed to hoarding them for future warfare, do not make them at all. The prioritization of offense over defense is a violation of the social contract between a state and its people (Brey, 2012).

5.2 Rule 4: The Sociotechnical Imperative

Rule 4 states that the design of computing artifacts must account for the sociotechnical systems in which they are embedded. Industrial Control Systems (ICS) are old, designed in the 1970s and 80s for a sociotechnical system where operators were inherently trusted (Knowles et al., 2015). They were not designed nor were they ever intended to connect to the modern internet. When connecting legacy systems to the web, one violates the Sociotechnical Imperative by embedding fragile systems in a hostile environment, like how in the Middle Ages forts would crumble under modern artillery fire. Therefore, Anticipatory ethicists can argue that critical infrastructure, specifically in power, water, and nuclear sectors must anticipate cyber warfare, designing systems

that have mechanical overrides and backups that run on a closed system. The reliance on insecure-by-design systems in important infrastructure is a failure of due diligence and a neglect of the moral duty to protect public safety (Miller et al., 2011).

6. Recommendations

Before outlining a normative framework, we must acknowledge the significant political and technical constraints hindering these solutions. State actors, particularly major cyber powers, benefit from the current ambiguity of the "Gray Zone." They have little incentive to sign treaties that would limit their strategic options. Technical challenges with attribution also make enforcing these norms difficult in real-time. The following recommendations offer an aspirational roadmap, a phased evolution of international law that begins with voluntary bilateral agreements before expanding into enforceable global treaties. Ambiguity must be resolved if we desire to prevent catastrophic escalation from cyber-attacks to cyber warfare to kinetic warfare. We propose a framework of Normative Constraints and Technical Defense that moves past the reactive model that is currently used.

The first recommendation is the creation of a Digital Geneva Convention (Microsoft, 2017). The international community needs to establish a treaty that clearly prohibits attacks targeting civilian infrastructure, regardless of the “Kinetic Equivalence.” Attacks targeting civilian infrastructure should be disallowed regardless of them being digital or physical. Attacks on the public sectors of the internet (like DNS root servers and BGP routing tables), healthcare systems, and nuclear safety systems must be designated as involving clear acts of war. This disposes of the ambiguous “effects-based” analysis. Under an effects-based analysis, if an attack on a hospital killed zero patients, it would be ignored, under a norms-based framework the act of attempting to target the hospital would be a war crime alone. This restores the Principle of Distinction by creating recognized safe zones in cyberspace that are off-limits to cyber-attacks and cyber conflicts (Finnemore & Hollis, 2016).

The second recommendation is implementing Mandatory Attribution Markers, acting like Shoulder Insignia. States must adopt standards where cyber operations performed by state actors require a cryptographic signature which identifies the state of origin. While this is counter intuitive to espionage, it follows the Geneva Convention standard of combatants wearing uniforms and carrying arms openly. If a state desires the legal protections of “Combatant Immunity” for its hackers, the hackers need to identify their country of origin or be tried as criminals in the affected state. This reduces the risk of misidentifying the aggressor, and in turn the additional problem of unidentifiable hackers striking the wrong target.

The third recommendation is to create Active Defense strategies. States must transition from passive defense (patching and firewalls) to active defense (hunting state sponsored hackers) (Denning, 2014). The U.S. Cyber Command strategy of ‘defending forward’ employs the tactic of disrupting attacks at the source before they can cause damage. In the realm of ethics, this lowers the risk of damage by neutralizing threats early, but must in turn be heavily regulated to prevent a “Preemptive War.” Active defense should be limited to disrupting the tool or infrastructure of the attacker and must not involve killing or harming the personnel of the adversaries’ host nation.

7. Future Work

Future research must address the role of Artificial Intelligence in reacting to warfare thresholds. Cyber operations occur at machine speed; humans are too slow to keep up.

Research is required into the ethics and issues related to Autonomous Responses to cyber-attacks. If an AI defense system catches an intrusion into a power grid and the signature matches that of a known destructive wiper, should it be allowed to counter-strike without human authorization? We anticipate a future where AI driven malware and AI driven defense will fight against one another, eventually escalating conflict to a level in which humans can no longer intervene. Future work needs to focus on designing physical failsafe’s for automated cyber conflicts that automatically pause hostilities when certain markers of damage or speed are passed (Taddeo & Floridi, 2018).

8. Conclusion

The threshold for determining when and how cyber-attacks rise to the level of cyber warfare persists as a dangerous ambiguity in geopolitics. The reliance on “Kinetic Equivalence” which is the idea that a cyber attack needs to look like an explosion to count at war is incredibly outdated. The case studies of Stuxnet, NotPetya, and Triton show that state actors are persistently towing the line. They exploit the Gray Zone to undermine

sovereignty and project power without having to declare war and face the consequences any nation conducting warfare would typically endure.

By applying Anticipatory Ethics, we come to the conclusion that the "Effects-Based" model is not sufficient for defining the threshold for defining when a cyber-attack rises to the level of cyber warfare because it relies on a post-hoc analysis. Waiting for damage to occur before calling something a cyber-attack or an act of cyberwarfare is a bad method of operating. To ensure a prosperous digital future, we must move towards a Norms-Based model which fundamentally moves the threshold for cyber warfare from consequentialism (damage-based judgement) to a deontological framework (judging an attack based on what its target is). The dissimilarity from the "Effects-Based" model, which requires a victim state to wait for kinetic equivalence, like physical destruction or the loss of life before any operation is considered an act of war. A Norms-Based model would create proactive "Red Lines" concerning categories of protected infrastructure. Within this framework the act of targeting systems that are indispensable to civilian survival, like hospital grids or election infrastructure, now becomes a violation of international norms and is classified as an act of aggression regardless of their success or lack thereof. This approach removes any ambiguity that the "Gray Zone" granted previous operations by focusing on intent as opposed to physical impact. Without a clear and enforceable definition of when a cyber-attack becomes an act of war and for what constitutes war in the cyber domain, we cannot hope for stable peace in the future.

Ethics Declaration: No human participants or personally identifiable information were involved. All data sources were publicly available.

AI Tools Declaration: ChatGPT 5.1 for drafting and refinement. Human authors verified all content. Gemini 3.0 pro used for aiding in sourcing.

References

Association for Computing Machinery. (2018). *ACM Code of Ethics and Professional Conduct*. https://www.acm.org/code-of-ethics

Bratman, Michael. (1999). Intentions, Plans and Practical Reason. Center for the Study of Language and Information.

Brey, P. A. E. (2012). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, 6(1), 1–13.

Clement, J. A. (2018). Cyber Warfare and the Gray Zone. *The Cyber Defense Review*, 3(2), 53–62.

Denning, D. E. (2014). The Ethics of Cyber Conflict. *The Handbook of Information and Computer Ethics*, 407–426. Wiley.

Greenberg, A. (2019). *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers*. Doubleday.

Greenberg, A. (2018). The Untold Story of the 2018 Olympics Cyberattack, the Most Deceptive Hack in History. *Wired*. https://www.wired.com/story/untold-story-2018-olympics-destroyer-cyberattack/

Johnson, B., Caban, D., Krotofil, M., Scali, D., Brubaker, N., & Glyer, C. (2017). *Attackers Deploy New ICS Attack Framework "TRITON" and Cause Operational Disruption to Critical Infrastructure*. FireEye Mandiant.

Kello, L. (2013). The Meaning of the Cyber Revolution: Perils to Theory and Statecraft. *International Security*, 38(2), 7–40.

Knowles, W., Prince, D., Hutchison, D., Disso, J. F. P., & Jones, K. (2015). A Survey of Cyber Security Management in Industrial Control Systems. *International Journal of Critical Infrastructure Protection*, 9, 52–80.

Lucas, G. R. (2017). *Ethics and Cyber Warfare: The Quest for Responsible Security in the Age of Digital Warfare*. Oxford University Press.

Microsoft. (2017). *A Digital Geneva Convention to protect cyberspace*.

Microsoft. (2020). *Deep Dive into the Solorigate Second-Stage Activation: From Sunburst to Teardrop*. Microsoft Security Response Center.

Miller, K.W., et al. (2011). Moral Responsibility for Computing Artifacts: 'The Rules'. *IT Professional*, 13(3), 57–59.

Rid, T., & Buchanan, B. (2015). Attributing Cyber Attacks. *Journal of Strategic Studies*, 38(1-2), 4–37.

Rid, T. (2012). Cyber War Will Not Take Place. *Journal of Strategic Studies*, 35(1), 5–32.

Schmitt, M. N. (Ed.). (2017). *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. Cambridge University Press.

Schmitt, M. N. (2021). SolarWinds and the International Law of Cyberspace. *The Cyber Defense Review*, 6(2), 29–46.

Sobczak, B. (2019). *The Inside Story of the World's Most Dangerous Cyberattack*. E&E News.

Taddeo, M., & Floridi, L. (2018). Regulate artificial intelligence to avert cyber arms race. *Nature*, 556(7701), 296–298.

Waxman, M. C. (2013). Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4). *Yale Journal of International Law*, 36, 421–459.

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