2020 12th International Conference on Cyber Conflict
20/20 Vision: The Next Decade
T. Jančárková, L. Lindström,
M. Signoretti, I. Tolga, G. Visky (Eds.)
2020 © NATO CCDCOE Publications, Tallinn

Retorsion as a Response to Ongoing Malign Cyber Operations

Jeff Kosseff
Assistant Professor
Cyber Science Department
United States Naval Academy[1]
Annapolis, MD
kosseff@usna.edu

Abstract: If a state has experienced a malicious cyber act that violates international law, it may implement proportional and limited countermeasures. If the act constitutes an armed attack, the target state may engage in self-defence. But what if the initial act, while malicious and harmful, does not clearly violate an international legal obligation? In such an instance, the primary option for response is retorsion, which is defined as an unfriendly but legal act. Little scholarship has meaningfully examined the contours of retorsion, which is increasingly important in an era of persistent, low-intensity cyber aggression. This paper seeks to fill that gap by exploring the contours of retorsion and examining the types of responses that could fall within its scope. It argues for an expansive view of retorsion that encompasses any responses that comport with international law. Definitional clarity is increasingly important to allow states to understand the range of potential responses to persistent cyber aggression that do not necessarily violate international law. Among the types of activities that may fall within the scope of retorsion are: exerting pressure via international relations, gathering information from the adversary's networks, observing the adversary on one's own network using tools such as honeypots, sending warnings to individual operatives, establishing a position on the adversary's systems and slowing down malign cyber operations.

Keywords: *retorsion, countermeasures, sovereignty, cyber*

1 The views expressed in this paper are only those of the author, and do not represent those of the United States Naval Academy, United States Department of Navy, United States Department of Defense, or any other party. Thanks to those who provided incredibly valuable feedback on earlier drafts, including Dennis Dias, Chris Inglis, Martin Libicki, and Kurt Sanger.

1. INTRODUCTION

In July 2019, the Netherlands Minister of Foreign Affairs released a nine-page summary of the government's views on international law as it applies to cyberspace. The document concluded with a discussion of states' response options, and much of the discussion focused on responses that have been thoroughly discussed in international law circles: countermeasures, pleas of necessity and self-defence. The document also highlighted a response that has not often been discussed in depth: retorsion.

As the Netherlands government described it, retorsion "relates to acts that, while unfriendly, are not in violation of international law".[2] Because retorsion is legal, the government wrote, it "is therefore always available to states that wish to respond to undesirable conduct by another state, because it is a lawful exercise of a state's sovereign powers."[3] The government listed a few examples: economic penalties, expelling diplomats, and "limiting or cutting off the other state's access to servers or other digital infrastructure in its territory".[4] Although the document only devoted two paragraphs to retorsion, the mere fact that a government highlighted it as one of the primary responses to malign cyber actions was noteworthy.

Retorsion is both flexible and limited. It is flexible because, unlike other responses, it is subject to relatively few operational requirements. It is limited because it may only consist of actions that comply with international law.

This paper highlights the reasons to classify a response to malign cyber activity as retorsion rather than countermeasures, and examines the types of responses that could qualify as retorsion. It argues for policymakers to broadly conceive of retorsion by including any responses – no matter how unfriendly – that comport with international legal norms, regardless of the legal status of the adversary's actions. A number of responses do not violate sovereignty or other international law. Very little scholarship has focused substantially on the boundaries of retorsion; this paper seeks to fill that gap.

A clearer understanding of retorsion is particularly useful as states confront persistent levels of cyber aggression that are below the level of an armed attack,[5] removing

2 Netherlands Minister of Foreign Affairs, Letter to the Parliament on the International Legal Order in Cyberspace (July 5, 2019), Appendix: International Law in Cyberspace, available at https://www.government.nl/binaries/government/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace/International+Law+in+the+Cyberdomain+-+Netherlands.pdf, at 7.

3 Ibid.

4 Ibid.

5 Statement of General Paul M. Nakasone, Commander, United States Cyber Command, Before the Senate Committee on Armed Services (Feb. 14, 2019), available at https://www.armed-services.senate.gov/imo/media/doc/Nakasone_02-14-19.pdf, at 2 ("The nation faces threats from a variety of malicious cyber actors, including non-state and criminal organisations, states, and their proxies. We see near-peer competitors conducting sustained campaigns below the level of armed conflict to erode American strength and gain strategic advantage").

self-defence as a potential response. Countermeasures will likely be available as a response to some of this aggression, but the use of countermeasures faces a number of constraints, described below. Many of the more aggressive responses will constitute countermeasures or even self-defence, but retorsion provides states with a flexible framework to respond to this persistent, low-level aggression.

2. THE LIMITS OF COUNTERMEASURES AS A LEGAL BASIS FOR RESPONSE

Countermeasures and retorsion are the primary legal categories of responses to cyber aggression that falls below the level of an armed attack. The main difference between them is that countermeasures would violate international law absent illegal actions by the adverse party,[6] while retorsion comports with international law regardless of the adverse party's actions.[7] Thus, outlining the limits that international law places on countermeasures helps to illustrate why it is useful to have a better understanding of the scope of retorsion.

Perhaps the most significant limit of countermeasures is that they only can be taken against a state that has violated an international legal obligation owed to the state seeking to take the countermeasure.[8] When it is debatable whether such a violation has occurred, there is uncertainty as to whether countermeasures are permissible. As one example, hacking that interferes with a state's electoral process could be viewed as a violation of the principle of non-intervention, which "forbids all States or groups of States to intervene directly or indirectly in internal or external affairs of other States",[9] but some commentators argue that election interference does not meet the standard for an illegal intervention.[10]

Often, such alleged violations that give rise to countermeasures involve breaches of sovereignty.[11] Determining whether such a breach has occurred in cyberspace is

6 Int'l Law Comm'n, Draft Articles on Responsibility of States for Internationally Wrongful Acts, Rep. of the Int'l Law Comm'n on the Work of Its Fifty-Third Session, U.N. Doc. A/56/10, at 75 (2001) [hereinafter ILC Draft Articles on Responsibility] at 75 ("In certain circumstances, the commission by one state of an internationally wrongful act may justify another state injured by that act in taking non-forcible countermeasures in order to procure its cessation and to achieve reparation for the injury").

7 Ibid. at 128.

8 Ibid. at 129 (Article 49.1) ("An injured State may only take countermeasures against a State which is responsible for an internationally wrongful act in order to induce that State to comply with its obligations under Part Two").

9 Craig Forcese, *The 'Hacked' US Election: Is International Law Silent, Faced with the Clatter of Cyrillic Keyboards*, JustSecurity (Dec. 16, 2016) , available at https://www.justsecurity.org/35652/hacked-election-international-law-silent-faced-clatter-cyrillic-keyboards (quoting *United States v. Nicaragua*).

10 Jens David Ohlin, *Did Russian Cyber Interference in the 2016 Election Violate International Law?* 95 Tex. L. Rev. 1579, 1587 (2017) (asserting that "the technical requirements for an illegal intervention might not apply to the Russian intervention, depending on how one understands the concept of coercion").

11 Michael N. Schmitt, *'Below the Threshold' Cyber Operations: The Countermeasures Response Option and International Law*, 54 Va. J. Intl'l L. 697, 704 (2014) ("In the cyber context, sovereignty grants a State the right (and in some cases the obligation) to regulate and control cyber activities and infrastructure on its territory").

often difficult, as there is no clear consensus as to whether an act of cyber aggression could constitute a standalone violation of sovereignty, or if it must implicate another rule such as non-intervention. The authors of the *Tallinn Manual* adopted the former view, writing that "[s]tates enjoy sovereignty over any cyber infrastructure located on their territory and activities associated with that cyber infrastructure".[12] Rule 4 of the *Tallinn Manual* provides that "[a] State must not conduct cyber operations that violate the sovereignty of another State".[13] This is consistent with the view of cyberspace sovereignty that the French Ministry of Armies released in 2019,[14] as well as that of some scholars.[15] In contrast, Jeremy Wright, then the Attorney General of the United Kingdom, said in 2018 that he was "not persuaded that we can currently extrapolate from that general principle a specific rule or additional prohibition for cyber activity beyond that of a prohibited intervention".[16] In other words, the general principle of sovereignty, in his view, did not create a bright-line rule that would be violated merely by virtue of an intrusion on the cyber infrastructure of another state. Similarly, an internal 2017 memorandum from the United States Department of Defense General Counsel stated that "[m]ilitary cyber activities that are neither a use of force, nor that violate the principle of non-intervention are largely unregulated by international law at this time".[17] Accordingly, states seeking to enact countermeasures may lack certainty that other states would view their actions as permissible responses.

In addition to the legal uncertainty over whether a particular act legally justifies countermeasures, the target state must have sufficient *factual* certainty of the source of the malign activity before engaging in countermeasures. As Michael Schmitt has noted, if its assessment of the origins of an attack "turns out not to be well-founded, the injured state's action cannot qualify as a countermeasure".[18] The Draft Articles on State Responsibility suggest "reasonable certainty" in attribution, leading Schmitt to conclude that "[a] cyber countermeasure undertaken in a mistaken but reasonable belief as to the identity of the originator or place of origin will be lawful so long as

12 *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, at 11 (2017) [hereinafter Tallinn Manual].

13 Ibid. at 17.

14 Przemyslaw Roguski, *France's Declaration on International Law in Cyberspace: The Law of Peacetime Cyber Operations, Part I*, Opinio Juris (Sept. 24, 2019) ("From this France concludes that any cyberattack, i.e. any operation which breaches the confidentiality, integrity or availability of the targeted system, constitutes at minimum a violation of French sovereignty, if attributable to another State.").

15 Sean Watts & Theodore Richard, *Baseline Territorial Sovereignty and Cyberspace*, 22 Lewis & Clark L. Rev. 803, 808 (2018) (arguing that "the baseline rules of territorial sovereignty should be currently understood as a rule of conduct that generally prohibits states' nonconsensual interference with the integrity of cyber infrastructure on the territory of other states").

16 Speech, Rt. Hon. Jeremy Wright, *Cyber and International Law in the 21st Century* (May 23, 2018) , available at https://www.gov.uk/government/speeches/cyber-and-international-law-in-the-21st-century ("Online as well as everywhere else, the principle of sovereignty should not be used by states to undermine fundamental rights and freedoms and the right balance must be struck between national security and the protection of privacy and human rights").

17 Watts & Richard (supra n 15) at 860 (quoting Memorandum from Jennifer M. O'Connor, Gen. Counsel of the Dep't of Def., International Law Framework for Employing Cyber Capabilities in Military Operations (Jan. 19, 2017)).

18 Schmitt (supra n 11) at 726.

all other requirements for countermeasures have been met".[19] Even under a flexible standard of "reasonable certainty," it may be difficult to attribute a malign act to a particular state.[20]

If a target nation identifies a violation of an international obligation and attributes it to another state with sufficient certainty, that state may engage in countermeasures; however, these countermeasures must be limited in purpose. Article 49 of the Draft Articles on Responsibility states that countermeasures may only be taken "to induce that state to comply with its obligations".[21] The purpose of limiting countermeasures is to reduce the likelihood of conflict escalation.[22] Article 49 limits countermeasures "for the time being",[23] which the drafters stated is meant to indicate "the temporary or provisional character of countermeasures".[24] If the initial malign actions that triggered the countermeasures are no longer occurring, the target state may no longer have the authority to engage in the countermeasures. Determining when illegal behaviour has ceased is difficult in cyberspace, particularly in light of the barrage of threats that nations face, often from the same handful of bad actors.

In addition to limits on the purpose and duration of countermeasures, international law restricts their magnitude. Article 51 of the Draft Articles requires that countermeasures be proportional, which means that they "must be commensurate with the injury suffered, taking into account the gravity of the internationally wrongful act and the rights in question".[25] The drafters of the *Tallinn Manual* suggested that when states consider whether countermeasures are proportional, they should assess "the injury suffered (i.e., the extent of harm), the gravity of the wrongful act (i.e., the significance of the primary rule breached), the rights of the injured and responsible State (and interests of other States that are affected, and the need to effectively cause the responsible State to comply with its obligations".[26]

Imagine that State A's government computer systems were taken offline for a day by a DDOS attack originating in State B. Using a countermeasure, State A might seek to cause damage to the State B computers that executed the attack. It would not be proportional, however, for State A to disable the power grid in an entire metropolitan area within State B.

19 Ibid. at 727.
20 Matthew C. Waxman, *Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)*, 36 Yale J. Int'l L. 421, 443 (2011) ("As a technical matter, those who study the problem of legally regulating cyber-attacks are usually quick to point out the problems of identification and attribution: it is not always possible to discern quickly or accurately who launched or directed an attack").
21 ILC Draft Articles on Responsibility at 129.
22 *Tallinn Manual* at 116.
23 ILC Draft Articles on Responsibility at 129.
24 Ibid.
25 Ibid. at 134.
26 *Tallinn Manual* at 128.

Countermeasures also face procedural restrictions. Most notably, a state engaging in countermeasures must "notify the responsible State of any decision to take countermeasures and offer to negotiate with that State [… though] the injured State may take such urgent countermeasures as are necessary to preserve its rights".[27] The notification requirement would pose little problem if the countermeasure was intended to dissuade the responsible state from continuing its malign acts. However, if the countermeasure was a cyber operation targeting the responsible state's systems that were responsible for the acts, notification would likely undercut the efficacy of the operation by providing a warning.

International law is also unsettled as to whether non-injured states may collectively engage in countermeasures on behalf of other states that are injured. In 2019, President Kersti Kaljulaid of Estonia took the position that international law allows collective countermeasures,[28] but that position at the moment is not widely accepted.[29] Unless there is a stronger international consensus that collective countermeasures are permissible, states will likely lack the necessary assurances to collaborate on countermeasures on behalf of an injured state.

In sum, countermeasures can be a useful tool to respond to low-intensity aggression in cyberspace, but their implementation is subject to many constraints. As states look to respond to this persistent malicious behaviour, they should consider whether some responses can be classified as retorsion rather than countermeasures.

3. RETORSION'S FLEXIBILITY

Countermeasures are subject to numerous restrictions because, absent the illegal acts of the responsible state, they would violate international law. If the actions underlying the countermeasures would not violate international law regardless of the actions of the other state, then it is unnecessary to classify them as countermeasures. Legal actions are retorsion that are not subject to the same limits as countermeasures. This section outlines the scope of retorsion and argues that it allows for a flexible approach to any unfriendly actions that comply with international law.

27 ILC Draft Articles on Responsibility at 135.

28 *'President Kaljulaid at CyCon 2019: Cyber Attacks Should Not be an Easy Weapon'*, ERR NEWS (May, 29, 2019), https://news.err.ee/946827/president-kaljulaid-at-cycon-2019-cyber-attacks-should-not-be-easy-weapon ("Estonia is furthering the position that states which are not directly injured may apply countermeasures to support the state directly affected by the malicious cyber operation").

29 Michael Schmitt, *"France's Major Statement on International Law and Cyber: An Assessment"*, JustSecurity (Sept. 16, 2019), available at https://www.justsecurity.org/66194/frances-major-statement-on-international-law-and-cyber-an-assessment/ ("Somewhat surprisingly in light of its central place in the NATO alliance and its key role in European security affairs, France rejects the position recently set forth by Estonian President Kersti Kaljulaid that collective countermeasures – that is, countermeasures taken by one State on behalf of another State that is entitled to take countermeasures by virtue of being the target of an unlawful cyber operation – are permissible").

Only a limited body of scholarship and jurisprudence has attempted to define retorsion, and often in a fleeting manner. The Draft Articles commentary describes retorsion as "'unfriendly' conduct which is not inconsistent with any international obligation of the State engaging in it even though it may be a response to an internationally wrongful act".[30] Likewise, the US Defense Department *Law of War Manual* characterises retorsion as "unfriendly conduct, (1) which is not inconsistent with any international obligation of the State engaging in it, and (2) which is done in response to an internationally wrongful act".[31] Leading treatises[32] and scholarship[33] similarly describe retorsion as unfriendly but legal actions. The scholarship typically focuses on diplomatic and economic forms of retorsion, such as sanctions.[34] However, if one were to take a broader view of retorsion so that it encompasses any unfriendly but legal response, these are but one form of retorsion.

A state that classifies its response as retorsion faces fewer legal constraints than if it employs countermeasures. The primary *legal* limit on retorsion is that, regardless of the actions of the adversary, it may not violate international legal obligations owed to other states. This eliminates a number of more aggressive options that may violate legal obligations,[35] but once a state has addressed the threshold concern of legality, it does not face as many legal restrictions on the purpose, duration, and character as a state employing countermeasures. Retorsion "casts a political shadow over the relationship between the two states", but such political effects are not legally prohibited.[36] Pragmatic concerns may limit retorsion, but such limits are not imposed by international law.

Unlike countermeasures, retorsion is not limited to responding to internationally wrongful acts. It may also be exercised in response to the unfriendly but legal acts

30 ILC Draft Articles on Responsibility at 128.

31 US Department of Defence, *Law of War Manual* (June 2015, Updated December 2016) at 1110 [hereinafter *Law of War Manual*].

32 Anthony Cassese, *International Law* (2d ed. 2005) at 310 ("Retortion embraces any retaliatory act by which a state responds, by an unfriendly act not amounting to a violation of international law, to either (a) a breach of international law or (b) an unfriendly act, by another state"); L. Oppenheim, *International Law: A Treatise* (1912) at 36-37 ("The act which calls for retaliation is not an illegal act; on the contrary, it is an act that is within the competence of the doer").

33 Catherine Lotrionte, *Reconsidering the Consequences for State-Sponsored Hostile Cyber Operations Under International Law*, Cyber Defence Review (2018) at 92 ("An act of retorsion is a coercive, politically unfriendly, but lawful act, not involving any breach of international obligations owed to the target state, whether treaty-based or customary and thereby do not require any legal justification"); Lindsay Moir, *The Implementation and Enforcement of the Laws of Non-International Armed Conflict*, 3 J. Armed Conflict L. 163, 176 (1998) ("Retorsion is an unfriendly, even potentially damaging, act. Unlike reprisals, however, retorsion is perfectly valid under international law"); Lori Fisler Damrosch, *Enforcing International Law Through Non-Forcible Measures* (1997) at 54 (defining "retorsion" as "an unfriendly (but not otherwise illegal) act taken in response to an unfriendly or illegal act").

34 Troy Anderson, "*Fitting a Virtual Peg into a Round Hole: Why Existing International Law Fails to Govern Cyber Reprisals*, 34 Aris. J. Int'l & Compl L. 135, 142 (2016) ("Retorsion usually is diplomatic or economic in nature, rather than militaristic").

35 Ibid. at 147 ("Limiting a cyber operation to the confines of legality in order to allow it to qualify as a legal retorsion severely limits the power of the cyber operation").

36 Bruno Simma (ed), *The Charter of the United Nations: A Commentary*, (1994) at 104.

of another state.[37] This allows states to develop responses to adverse actions without first engaging in a legal analysis of whether the adversary's actions violated an international legal obligation.

Unlike countermeasures, retorsion is not necessarily confined to a particular goal. Some commentators have suggested that, as with countermeasures, it is often intended to persuade a state to cease its internationally wrongful acts,[38] but there is nothing in the Draft Articles or other authoritative commentary that would confine retorsion to mere persuasion. It also could include measures that blunt the harm of an adversary's actions, or prevent the adversary from exercising its capabilities. Indeed, the legal nature of retorsion means that it is not subject to the same limitations as other responses.[39] Relatedly, if a state engages in retorsion rather than countermeasures, it does not have a legal obligation to notify the other state.

Retorsion is not subject to the strict duration requirements of other responses. As described in the previous section, countermeasures must cease immediately once the other state has ceased its internationally wrongful acts. Oppenheim suggested in 1912 that because "retorsion is made use of only to compel a state to alter its discourteous, unfriendly, or unfair behaviour, all acts of retorsion ought at once to cease when such State changes its behaviour".[40] Even under this limited view – which is not supported by more recent authoritative sources – retorsion need not cease once the other state is *legally* compliant; indeed, a legal violation is not a precondition for retorsion. Oppenheim's suggestion, to the extent that it is followed, is a more pragmatic and political guideline: for instance, if State A's sanctions caused State B to stop attacking State A's election system, then it would be politically and diplomatically unwise for State A to continue the sanctions unless there was an indication of further malicious action by State B.

Nor does international law require retorsion to be proportionate to the malign actions of the adversary, as is required for countermeasures. *Brierly's Law of Nations* notes that "it is sometimes suggested that retaliation should be proportionate",[41] but it cites no binding or persuasive legal precedent that would suggest proportionality is a requirement for retorsion. As with the limit on duration, it may be that a disproportionate retorsion would raise political or diplomatic concerns, but proportionality is not a legal requirement of retorsion, which by definition is an independent legal act.

37 Andrew Clapham, *Brierly's Law of Nations* (7th ed. 2012) at 397 ("'Retorsion' is a measure of self-help taken in response to an illegal or unfriendly act, where the self-help measure itself is within the law").

38 Edward Kwakwa, *Belligerent Reprisals in the Law of Armed Conflict*, 27 Stanford J. Int'l L. 49, 51 (1990) ("a retorsion seeks to coerce another state to discontinue a vexatious or injurious – but legal – practice").

39 *Law of War Manual* (supra n 31) at 1110 ("Because retorsion, by definition, does not involve the resort to actions that would ordinarily be characterised as illegal, the stringent conditions that apply to reprisal do not apply to retorsion").

40 Oppenheim (supra n 31) at 38.

41 Clapham (supra n 37) at 397.

Unlike countermeasures, international law does not restrict nations from collaborating on retorsion. For instance, if a state has repeatedly acted maliciously in cyberspace with targets in multiple states, all of those states could collectively engage in sanctions or release a joint public statement condemning the bad actor.

In sum, retorsion is both narrow and broad. It is narrow in the sense that it only applies to actions that, standing independently, would not violate international law. It is broad because, if the acts qualify as retorsion, they are not subject to the same legal constraints as responses such as countermeasures. Retorsion is often overlooked in debates on international law, which tend to focus on countermeasures and self-defence. While those frameworks are vital to the discussion, we also must examine whether responses can constitute retorsion and are afforded more leeway.

4. ACTIONS THAT MAY CONSTITUTE RETORSION IN RESPONSE TO MALIGN CYBER OPERATIONS

Whether a response qualifies as retorsion depends entirely on whether the measure violates any international legal rules. Some actions such as intentionally causing damage to another state's computer systems do not constitute retorsion because they likely violate a legal obligation. But what *does* qualify as retorsion in the cyber realm? Legal scholarship often provides sanctions as the primary example of retorsion. Based on the definition of retorsion set out in this paper, sanctions in response to malign cyber actions clearly would qualify as retorsion. However, this paper posits that sanctions are only one form of retorsion, and policymakers should search more broadly for responses to cyber actions that are legal and therefore not subject to the same restrictions as countermeasures. This section categorises the types of responses that might fit into the broader concept of retorsion. To assess whether these actions qualify as retorsion, it is necessary to determine whether they violate sovereignty, prohibitions on the use of force or other legal norms.

A. Pressure via International Relations

The classic and most oft-cited examples of retorsion involve a target state using standard tools of international relations to pressure an adversary to stop its illegal or unfriendly acts. Such examples include "severance of diplomatic relations and the expulsion or restrictive control of aliens, as well as various economic and travel restrictions".[42] So too is a US law requiring suspension of foreign aid "to any country nationalising American property without proper compensation",[43] and the April 2015 executive order that allows sanctions for, among other things, "harming, or otherwise significantly compromising the provision of services by, a computer or network of

42 Malcom N. Shaw, *International Law* (8th ed. 2017) at 859.

computers that support one or more entities in a critical infrastructure sector".[44] In 2019, the US imposed sanctions on North Korean hackers accused of a number of cyber operations, including the 2014 hack of Sony Pictures.[45] Likewise, in December 2016, the US expelled 35 Russian diplomats in response to Russian interference in the 2016 US elections.[46]

A target state could also publicly shame a nation that has attacked its systems. For instance, countries are increasingly securing indictments in their domestic courts against foreign hackers.[47] In many cases, the countries issuing the indictments do not have extradition arrangements with the states where the hackers are located, so it is unlikely that the hackers will ever stand trial. However, the indictments play an important role in publicly "naming and shaming" both the individual cyber operators and, in many cases, the governments that employ them.

These responses are all, to varying degrees, unfriendly; yet they do not violate any international legal principles and therefore clearly qualify as retorsion. They do, however, face a number of political constraints. For example, if State A mistakenly attributes a DDOS attack to State B and implements sanctions, it risks significant diplomatic pushback from State B and other states. However, such a short-sighted act would not violate international law and is not subject to the same limits as countermeasures.

B. Accessing Information on the Adversary's Systems

A state that has been targeted by another state's hostile cyber acts may seek to access that state's systems to gather information about the adversary's operations. To the extent that this constitutes legal peacetime espionage, it should fall under the broad umbrella of retorsion. The general rule is that peacetime cyber espionage is not illegal per se.[48] Of course, a state still could violate another state's sovereignty if, for instance, an act of espionage causes damage to data or computer systems.[49] Moreover, although the prevailing view, as stated in the *Tallinn Manual*, is that peacetime cyber espionage

43 Ibid.

44 Executive Order 13694 (April 1, 2015).

45 Carol Morello & Ellen Nakashima, *US Imposes Sanctions on North Korean Hackers Accused in Sony Attack, Dozens of Other Incidents*, Wash. Post (Sept. 13, 2019), available at https://www.washingtonpost.com/national-security/us-sanctions-north-korean-hackers-accused-in-sony-attack-dozens-of-other-incidents/2019/09/13/ac6b0070-d633-11e9-9610-fb56c5522e1c_story.html.

46 David E. Sanger, *Obama Strikes Back at Russia for Election Hacking*, NY Times (Dec. 29, 2016), available at https://www.nytimes.com/2016/12/29/us/politics/russia-election-hacking-sanctions.html.

47 Alfred Ng, 'Justice Department Charges North Korean Over WannaCry, Sony Hack', CNET (Sept. 6, 2018), available at https://www.cnet.com/news/justice-department-charges-north-korean-hacker-linked-to-wannacry-2014-sony-hack/.

48 *Law of War Manual* at 1016 ("Generally, to the extent that cyber operations resemble traditional intelligence and counter-intelligence activities, such as unauthorised intrusions into computer networks solely to acquire information, then such cyber operations would likely be treated similarly under international law").

49 *Tallinn Manual* at 170 ("For instance, if organs of one State, in order to extract data, hack into the cyber infrastructure located in another State in a manner that results in a loss of functionality, the cyber espionage operation violates, in the view of the Experts, the sovereignty of the latter").

per se is legal, some governments have pushed back against this rule and argued that in some cases espionage may be an infringement of sovereignty.[50]

The current majority view, however, is that unless cyber espionage results in damage, it does not violate international law. To the extent that this remains the general rule, conducting espionage operations on an adversary's systems may constitute retorsion. For instance, assume that State A has repeatedly attempted to spread false information to interfere in the elections of State B. State B may attempt to access data on State A's computers that provides insight into this propaganda campaign. Assuming that this stays within the boundaries of legal peacetime cyber espionage, State B need not attempt to classify its action as a countermeasure, as it would constitute retorsion. Such characterisation is particularly useful in this scenario, as there is considerable debate as to whether such election interference constitutes a breach of international legal obligations. To the extent that State B's hacking operations can be characterised as retorsion, it need not concern itself with whether State A's campaign was legal.

The limited commentary about retorsion does not include espionage among the examples of retorsion, as the commentary typically focuses on responses such as sanctions and expulsion of diplomats. However, if we are to view retorsion as any unfriendly act that complies with international law, many forms of espionage would also fall within the definition of retorsion.

C. Conducting Cyber Operations on One's Own Network (Honeypots and Sinkholes)

The adversary's systems are not the only potential source of information about their capabilities and plans. A target state could learn about the adversary by observing their actions on the target state's own systems. Such operations are even more likely to qualify as retorsion than the espionage described above. Unlike the operations that take place on the adversary's systems, such local observations do not even come close to raising any questions of territorial sovereignty violations.

What if the target state took steps to entice the adversary to be present on its network, allowing the target state to observe the adversary's actions? For instance, imagine that State A has not only been launching propaganda to influence State B's elections, but also attempting to access and delete voting data from State B's elections systems. State B might use a honeypot[51] to lure State A to a particular State B server, and then observe State A's actions and gather information about its techniques. Honeypots

50 Russel Buchan, *The International Legal Regulation of State-Sponsored Cyber Espionage*,in International Cyber Norms: Legal, Policy & Industry Perspectives, Anna-Maria Osula, Henry Rõigas (eds.) (2016) at 71 ("There is state practice to suggest that where a state considers itself to have been the victim of cyber espionage it regards such behaviour as falling foul of the principle of territorial sovereignty").

51 Paul Rosenzweig, *International Law and Private Actor Active Cyber Defensive Measures*, 50 Stan. J. Int'l L. 103, 106 n8 (2014). ("As the name implies, honeypots are intended to attract hackers by purporting to be worthwhile subjects of attack. One might, for example, give a document honeypot the Microsoft Word name 'Plans for Countering Hackers.Docx' and expect it to be the subject of an attack").

can be quite useful both for obtaining information and distracting attackers on false systems.[52] Governments use two primary types of honeypots: production honeypots detect imminent threats and are easier to deploy, while research honeypots gather information about emerging tactics of adversaries.[53] Another tool, the sinkhole, diverts harmful traffic, such as a botnet, to prevent harm.[54]

This use of honeypots and sinkholes should qualify as retorsion because they do not involve the infringement of sovereignty or any other legal obligation to State A; indeed, the act takes place entirely on the systems of State B, once State A has accessed the system. Critics of this approach might argue that the use of such tools to deceive another state is more aggressive than traditional espionage. While this may be true, there is little support for a claim that such deception – occurring entirely on State B's systems as a result of State A's intentionally malicious actions – would violate international law. If, however, State B were to use data collected via the honeypot to cause damage to State A's systems, the act would probably no longer qualify as retorsion. State A could claim that State B caused damage by unnecessarily consuming State A's resources with a sinkhole, but that argument would not be likely to prevail because the distraction merely prevented State A from malicious acts against State B.

What if a state were to install malware in data exfiltrated from its network? Whether such an act would constitute retorsion would depend on the effects of the malware. If the malware merely allowed the target state to observe data on the adversary's network, such an action probably would constitute retorsion, as the impacts would be no different from other forms of espionage. However, if the malware caused damage to the adversary's systems, the act might not be classifiable as retorsion because it might violate the adversary's sovereignty, in which case it would need another justification such as countermeasures. The exact boundaries as to when honeypots constitute an internationally wrongful act are subject to significant debate.[55]

D. Influencing Adversaries

A state that has been targeted by state-sponsored hackers may seek to send a message to those hackers to discourage them from engaging in further such acts. Influence is one of the three primary operational components of the US Defense Department's

52 Ian Walden & Anne Flanagan, *Honeypots: A Sticky Legal Landscape*, 29 Rutgers Computer & Tech. L.J. 317, 319 (2003) ("It can serve as a decoy to deflect the hacker from breaking into the real system, as a research tool for systems administrators merely to observe and learn how hackers operate and about weaknesses in their systems, or as a tool to monitor and document evidence for criminal prosecution").

53 Josh Fruhlinger, *What is a Honeypot? A Trap for Catching Hackers in the Act*, CSO (April 1, 2019).

54 Lily Hay Newman, *Hacker Lexicon: What Is Sinkholing?* Wired (Jan. 2, 2010), available at https://www.wired.com/story/what-is-sinkholing/.

55 David Wallace & Mark Visger, *The Use of Weaponised 'Honeypots' Under the Customary International Law of State Responsibility*, Cyber Defence Review (Summer 2018) at 38 ("Moreover, is it not reasonable for a State defending its cyber infrastructure to take measures, like using honeypots, to protect itself against such intrusions and, quite frankly, deter others? Is it wrong for a State to use a dynamic, penalty-based form of deterrence? The law, as it is currently structured, does not address these questions").

operational concept of "Defend Forward".[56] For instance, in October 2018, the *New York Times* reported that US Cyber Command sent messages to Russian operatives who disseminated propaganda during US elections "telling them that American operatives have identified them and are tracking their work".[57]

Retorsion would be a particularly attractive classification for such an operation, as it would avoid the need to determine whether the Russian election propaganda constituted a breach of international law that justified countermeasures. Of course, the potential barrier to the retorsion classification would be a claim that the US messaging violated international law. As applied to the public reports of the Cyber Command operation, such an argument against retorsion is unlikely to succeed. The *Times* quoted senior defence officials anonymously stating that "they were not directly threatening the operatives",[58] so the operation likely does not raise any concerns about violating international humanitarian law.

A warning accompanied by a specific threat of physical injury to the hackers could violate sovereignty, prohibitions on threats of use of force and even international humanitarian law, but the target state could claim retorsion for a narrowly tailored message that simply makes the adversary aware the target is watching their actions. Such an action certainly is unfriendly, but it does not violate international legal obligations.

E. Establishing a Position on the Adversary's Network

If a state has been the target of malign cyber operations, it may seek to establish a position on the adversary's systems. Positioning, like influence, is a component of the US Defend Forward operational concept.[59] Such positioning serves two primary purposes. First, it might send a message to the adversary that further actions could have consequences. Robert Chesney has described such a move as a "hold at risk" operation, with the goal "of establishing access to a potential adversary's system is to bolster one's deterrence posture by making clear to the adversary that you are capable, as a practical matter, of overcoming their defences and harming something they value".[60] Second, establishing a position allows the target state to respond more

56 Jeff Kosseff, *The Contours of 'Defend Forward' Under International Law*, paper for the 2019 11th International Conference on Cyber Conflict (2019) at 5 ("The Defend Forward concept also encourages stability by disabusing adversaries of the idea that they can operate with impunity in cyberspace and signals US commitment to confront hostile activities and impose cumulative costs for ongoing malicious actions") (internal quotation marks and citations omitted).

57 Julian E. Barnes, *US Begins First Cyberoperation Against Russia Aimed at Protecting Elections*, NY Times (Oct. 23, 2018) available at https://www.nytimes.com/2018/10/23/us/politics/russian-hacking-usa-cyber-command.html.

58 Ibid.

59 Kosseff (supra n 56) at 5 "'Perhaps the biggest shift in US cyber operations under Defend Forward is Cyber Command's recognition of the need for a forward cyber posture that can be leveraged to persistently degrade the effectiveness of adversary capabilities and blunt their actions and operations before they reach US networks") (internal quotation marks and citation omitted).

60 Robert Chesney, *The 2018 DOD Cyber Strategy: Understanding 'Defense Forward' in Light of the NDAA and PPD-20 Changes*, Lawfare (Sept. 25, 2018), available at https://www.lawfareblog.com/2018-dod-cyber-strategy-understanding-defense-forward-light-ndaa-and-ppd-20-changes.

quickly to any further harmful cyber actions by the adversary, perhaps allowing it to disable the source of the malign actions. Chesney refers to this as a "preparation of the battlefield" operation.[61]

There is a strong argument that merely accessing the adversary's systems – either to hold at risk or to prepare the battlefield – does not constitute a wrongful act under international law and therefore can be categorised as retorsion. Even if one were to recognise a standalone sovereignty obligation that is separate from other rules such as non-intervention, it is far from certain that mere access would violate that obligation. To be sure, if the target state were to leverage that access, such as by causing harm to the adversary's system, such an action might raise sovereignty concerns and not be justifiable as retorsion. Accordingly, it is important to separate the legal analysis of establishing a position on a network from the analysis of using that position.

F. Slowing Down the Adversary

Cyber operations may also attempt to impede the progress of an adversary who has conducted malign cyber operations. While it might be possible to classify such operations as countermeasures, there is at least a reasonable chance that the actions do not violate international legal obligations and therefore constitute retorsion. Consider Operation Glowing Symphony, a 2016 operation in which US Cyber Command accessed ISIS media systems, "deleted files, closed accounts, changed passwords", and "began moving through the ISIS networks they had mapped for months like a raid team clearing a house".[62] The operation resulted in ISIS media operatives being locked out of their accounts, having slow connections and other glitches.[63] Would such actions be permissible if conducted against a nation-state? Even under the expansive view of territorial sovereignty, it is at least debatable whether such inconveniences amount to a violation of international law. Under the views articulated in the 2017 US Department of Defense internal memorandum, cyber operations are only constrained by the prohibitions on the use of force and on intervention, and therefore there is an even stronger argument that Operation Glowing Symphony complied with international legal obligations. Indeed, some commentators have speculated that the release of the Defense Department memo soon after disclosure of Operation Glowing Symphony "raises the possibility it was produced to instruct DoD components of the legal analysis that supported the operation".[64] Such slow-down operations are unfriendly, but absent more significant harms there is at least a reasonable argument that the operations are retorsion.

61 Ibid.

62 Dina Temple-Raston, How the US Cracked Into One of the Most Secretive Terrorist Organisations, NPR (Sept. 26, 2019), available at https://choice.npr.org/index.html?origin=https://www.npr.org/2019/09/26/764790682/how-the-u-s-cracked-into-one-of-the-most-secretive-terrorist-organizations?t=1588800161071.

63 Ibid.

64 Watts & Richard (supra n 15) at 862.

5. CONCLUSION

This paper has sought to better define retorsion in an effort to provide states with more certainty about their options if countermeasures or self-defence are impractical or unavailable. Retorsion is typically associated only with international affairs such as sanctions and public denunciation. While those are critical examples of retorsion, other responses should fall under the same umbrella. All legal responses should be available for states to impede and discourage malign cyber actions. Before proceeding to an analysis of how to qualify a response as a countermeasure, a state should first determine whether it can justify its response as retorsion.