INTERNATIONAL LAW STUDIES

Published Since 1895

Adverse Cyber Operations: Causality, Attribution, Evidence, and Due Diligence

Hans-Georg Dederer and Tassilo Singer

95 INT'L L. STUD. 430 (2019)

Volume 95 2019

Stockton Center for International Law, U.S. Naval War College

ISSN 2375-2831

Adverse Cyber Operations: Causality, Attribution, Evidence, and Due Diligence

Hans-Georg Dederer and Tassilo Singer***

CONTENTS

* Full Professor, Dr. iur., Faculty of Law, University of Passau, Germany.

** Lecturer (2012–17), Dr. iur., Faculty of Law, University of Passau, Germany.

The authors would sincerely like to thank the Editor-in-Chief John Hursh for his excellent work in reviewing and editing this article.

The thoughts and opinions expressed are those of the authors and not necessarily those of the U.S. government, the U.S. Department of the Navy, or the U.S. Naval War College.

I. INTRODUCTION

Digitalization has driven the interconnectedness and interdependence of the world in an unprecedented way, especially through the creation and expansion of cyberspace.[1] Unsurprisingly, as with almost any highly innovative technological advances, the manifest and enormous benefits of digital globalization are accompanied by serious risks and unforeseen challenges.

One of the key challenges is adverse cyber operations against States, operations that are already on the rise and that will certainly escalate in the near future.[2] What appears to make adverse cyber operations so inviting is, for technical reasons, such operations are extremely difficult—if even possible—to attribute to a particular person, group of persons, or entity.[3] As long

1. *See, e.g.*, KLAUS W. GREWLICH, GOVERNANCE IN "CYBERSPACE": ACCESS AND PUBLIC INTEREST IN GLOBAL COMMUNICATIONS (1999). For an analysis of various legal topics concerning cyberspace and cyber activities, see RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE (Nicholas Tsagourias & Russell Buchan eds., 2015).

2. For a German perspective, see FEDERAL OFFICE FOR INFORMATION SECURITY, THE STATE OF IT SECURITY IN GERMANY 75 (2017); BUNDESMINISTERIUM DES INNERN, CYBER-SICHERHEITSSTRATEGIE FÜR DEUTSCHLAND [FEDERAL MINISTRY OF THE INTERIOR, CYBERSECURITY STRATEGY FOR GERMANY] 7 (2016). Among the most serious recent incidents are intrusions by alleged Russian hackers into the networks of U.S. energy suppliers that could have led to large-scale blackouts and an espionage operation against the German federal government's network also apparently attributable to a Russian group of hackers. *See Russian Hackers Penetrated Networks of U.S. Electric Utilities: WSJ*, REUTERS (July 28, 2018), https://www.reuters.com/article/us-usa-cyber-russia/russian-hackers-penetrated-networks-of-u-s-electric-utilities-wsj-idUSKBN1KE03F; *Bundesregierung: Hackerangriff auf Regierungsnetz "isoliert und unter Kontrolle"* [*Federal Government: Hacker Attack on Government Network "Isolated and under Control"*], HEISE ONLINE (Mar. 1, 2018), https://www.heise.de/newsticker/meldung/Bundesregierung-Hackerangriff-auf-Regierungsnetz-isoliert-und-unter-Kontrolle-3983757.html; *Bundesregierung wurde Gehackt* [*Federal Government was Hacked*], TAGESSCHAU (Feb. 28, 2018), https://www.tagesschau.de/inland/hackerangriff-regierungsnetz-101.htm.

3. Tracing cyber operations to a particular "person," "group of persons," or "entity" is essential because the rules on attribution of the Draft Articles on Responsibility of States for Internationally Wrongful Acts state that only the "conduct" of a "person," "group of persons," or "entity" can be attributed to a State, triggering the international responsibility of that State if the conduct breaches an international obligation. *See International Law Commission, Draft Articles on Responsibility of States for Internationally Wrongful Acts* arts. 2, 4, 5, 8, 56 U.N. GAOR Supp. No. 10, at 26, U.N. Doc. A/56/10 (2001), *reprinted in* [2001] 2 Yearbook of the International Law Commission 32, U.N. Doc. A/CN.4/SER.A/2001/Add.1 (Part 2)

as the causal links between the cyber operation and its adverse effects and the conduct of those responsible cannot be established precisely, attribution of the cyber operation to a particular State is not possible. Without accountability, or the possibility of being held accountable, adverse cyber operations are, and will remain, a very attractive tool for States to harm and destabilize other States.[4]

Besides causality and attribution problems, adverse cyber operations raise peculiar legal issues in public international law, such as classification. These include *jus ad bellum* questions, such as whether cyber operations constitute "armed attacks" within the meaning of Article 51(1) of the U.N. Charter and "use of force" within the meaning of Article 2(4). They also include *jus in bello* questions, such as whether cyber operations may give rise to an "armed conflict" triggering the applicability of international humanitarian law[5] or constitute an "attack" within the meaning of Article 49(1) of Additional Protocol I.[6] These issues have been dealt with comprehensively, most notably by the international group of experts that prepared *Tallinn Manual 2.0*.[7] Accordingly, this article will not take up these threshold issues.

[hereinafter ILC Draft Articles with Commentary]. For the customary international law status of the ILC Draft Articles on attribution, see Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro), Judgment, 2007 I.C.J. Rep. 43, ¶ 401 (Feb. 26).

4. *See*, *e.g.*, SVEN-HENDRIK SCHULZE, CYBER-"WAR" – TESTFALL DER STAATENVERANTWORTLICHKEIT [CYBER-"WAR" – A TEST CASE OF STATE RESPONSIBILITY] 130–31 (2015).

5. Convention (I) for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field art. 2, Aug. 12, 1949, 6 U.S.T. 3114, 75 U.N.T.S. 31; Convention (II) for the Amelioration of the Condition of the Wounded, Sick, and Shipwrecked Members of Armed Forces at Sea art. 2, Aug. 12, 1949, 6 U.S.T. 3217, 75 U.N.T.S. 85; Convention (III) Relative to the Treatment of Prisoners of War art. 2, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135; Convention (IV) Relative to the Protection of Civilian Persons in Time of War art. 2, Aug. 12, 1949, 6 U.S.T. 3516, 75 U.N.T.S. 287.

6. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3.

7. TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS (Michael N. Schmitt ed., 2d ed. 2017) [hereinafter TALLINN MANUAL 2.0]. With regard to the *jus ad bellum*, see, in particular, the *Manual*'s discussion of "use of force," *id.* at 330–37; "self-defense against armed attack," *id.* at 339–48. As concerns the *jus in bello*, see "international armed conflict," *id.* at 379–85; "non-international armed conflict," *id.* at 385–91; "cyber attack," *id.* at 415–20; *see also* Louisa Arimatsu, *Classifying Cyber Warfare*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 326, 328–32; William C. Banks, *Developing Norms for Cyber Conflict*, *in* RESEARCH HANDBOOK OF

Rather, this article will focus on those causality and attribution problems briefly outlined above. These problems form the focal points of the debate concerning responses to adverse cyber operations. Both causality and attribution lead to evidentiary questions, such as which party bears the burden of proof and what the applicable standard of proof is when a party alleges a breach of an international obligation. This article will address these questions and related evidentiary questions. We argue that, with regard to adverse cyber operations, causality, attribution, and evidentiary issues are also, and decisively, informed by the exercise of due diligence (or lack thereof) by the State of origin, that is, the State from whose territory the adverse cyber operation originates.[8] Accordingly, in Part III, we address due diligence requirements for States when cyberspace activities occur within their territory.[9]

REMOTE WARFARE 273, 277–87 (Jens David Ohlin ed., 2017); Yoram Dinstein, *Computer Network Attacks and Self-Defense*, 76 INTERNATIONAL LAW STUDIES 99, 102–03 (2002); Carlo Focarelli, *Self-Defence in Cyberspace*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 255, 263–70; Robin Geiß & Henning Lahmann, *Freedom and Security in Cyberspace: Shifting the Focus away from Military Responses towards Non-Forcible Countermeasures and Collective Threat-Prevention*, *in* PEACETIME REGIME FOR STATE ACTIVITIES IN CYBERSPACE: INTERNATIONAL LAW, INTERNATIONAL RELATIONS AND DIPLOMACY 621, 621–22 (Katharina Ziolkowski ed., 2013); Terry D. Gill, *International Humanitarian Law Applied to Cyber-Warfare: Precautions, Proportionality and the Notion of 'Attack' under the Humanitarian Law of Armed Conflict*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 366, 367–374; Wolff Heintschel von Heinegg, *Cyberspace – Ein völkerrechtliches Niemandsland?* [*Cyberspace – An International No Man's Land?*], *in* AUTOMATISIERUNG UND DIGITALISIERUNG DES KRIEGES [AUTOMATION AND DIGITIZATION OF WAR] 159, 162 (Roman Schmidt-Radefeldt & Christine Meissler eds., 2012); Marco Roscini, *Evidentiary Issues in International Disputes Related to State Responsibility for Cyber Operations*, 50 TEXAS INTERNATIONAL LAW JOURNAL 233, 239–40 (2015); Marco Roscini, *Cyber Operations as a Use of Force*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 233, 235–240; SCHULZE, *supra* note 4, at 83–84; Matthew J. Sklerov, *Solving the Dilemma of State Responses to Cyberattacks: A Justification for the Use of Active Defenses against States Who Neglect Their Duty To Prevent*, 201 MILITARY LAW REVIEW 1, 50–59 (2009); GARY D. SOLIS, THE LAW OF ARMED CONFLICT 679–86 (2d ed. 2016).

8. In this article, we will not argue that the violation of the due diligence obligation to prevent significant transboundary harm caused by certain conduct within a State's territory would imply, per se, that the conduct is attributable to that State. For such an approach, see, for example, Sklerov, *supra* note 7, at 60–62.

9. TALLINN MANUAL 2.0, *supra* note 7, at 30–50.

II. ADVERSE CYBER OPERATIONS AND POSSIBLE RESPONSES

Before delving into an analysis of these legal issues, the term "adverse cyber operation" requires definition. The term "cyber operations" and closely related terms have been defined in the *Tallinn Manual* as well as other academic works.[10] According to the *Manual*, a cyber operation is "the employment of cyber capabilities with the primary purpose of achieving objectives in or by the use of cyberspace."[11] Cyber operations include "cyber attacks" and "cyber exploitations." Cyber attacks are those operations "reasonably expected to cause injury or death to persons or damage or destruction to objects."[12] In contrast, cyber exploitations are operations where the sole focus is to "obtain information without affecting the functionality of the accessed system."[13] Finally, the term "cyber incident" frequently is used synonymously with cyber operations.[14]

In this article, we use the definition of cyber operation assigned by the *Tallinn Manual*.[15] Accordingly, an adverse cyber operation is the employment of cyber capabilities with the purpose of causing harm in or by the use of cyberspace.[16]

Causality, attribution, and evidentiary issues resulting from adverse cyber operations are invariably linked to the legal basis for the potential responses by States affected by the operation (the victim-State). These responses could

---

10. *See, e.g.*, Oliver Dörr, *Obligations of the State of Origin of a Cyber Security Incident*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 87, 88 (2015); Paul Ducheine, *The Notion of Cyber Operations*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 211. See also the legal definitions provided by the Convention on Cybercrime, ETS No. 185 (Nov. 23, 2001), https://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185.

11. TALLINN MANUAL 2.0, *supra* note 7, at 564; *see also* Roscini, *Cyber Operations as a Use of Force*, *supra* note 7, at 234.

12. TALLINN MANUAL 2.0, *supra* note 7, r. 92, at 415. For a more detailed definition, see Roscini, *Cyber Operations as a Use of Force*, *supra* note 7, at 234; *see also* WILLIAM H. BOOTHBY, WEAPONS AND THE LAW OF ARMED CONFLICT 239–40 (2d ed. 2016).

13. Roscini, *Cyber Operations as a Use of Force*, *supra* note 7, at 234.

14. *See, e.g.*, Christian Walter, *Cyber Security als Herausforderung für das Völkerrecht* [*Cyber Security as a Challenge to International Law*], 70 JURISTENZEITUNG 685, 686–87 (2015) [hereinafter Walter, *Cyber Security*].

15. *See supra* note 11 and accompanying text.

16. Concerning the various forms of adverse cyber operations, see, for example, SCHULZE, *supra* note 4, at 24.

flow from the right to self-defense under Article 51(1) of the U.N. Charter[17] or, arguably, the right to adopt countermeasures.[18] Indeed, some scholars conclude that countermeasures against adverse cyber operations may take the form of so-called "active defenses" or "active cyber defenses,"[19] which are "in-kind response(s) . . . against the attacker's system."[20] These measures, also called "hack backs,"[21] are directed to stop the attack by disabling its source.[22] Finally, the victim-State may put forward a "plea of necessity"[23] to justify an otherwise unlawful response to an adverse cyber operation. However, the International Court of Justice (ICJ) has held that a "state of necessity" can be invoked only under very restrictive circumstances.[24]

III. CAUSALITY AND ATTRIBUTION

Regardless of whether the response is premised on the right of self-defense, the right to adopt countermeasures, or a plea of necessity, the causal links between the cyber operation and its adverse effects, and the person, group of persons, or entity who conducted the cyber operation must be established. In addition, it may be necessary to meet certain factual predicates to attribute the perpetrator's conduct to a particular State.

---

17. *See, e.g.*, Dinstein, *supra* note 7, at 99–102.

18. *See* ILC Draft Articles with Commentary, *supra* note 3, arts. 22, 49–53.

19. For a definition of active cyber defense, see TALLINN MANUAL 2.0, *supra* note 7, at 563.

20. For this definition of active defenses, see Sklerov, *supra* note 7, at 25.

21. *See, e.g.*, Jan E. Messerschmidt, *Hackback: Permitting Retaliatory Hacking by Non-State Actors as Proportionate Countermeasures to Transboundary Cyberharm*, 52 COLUMBIA JOURNAL OF TRANSNATIONAL LAW 275 (2013). For one definition of a hack back, see TALLINN MANUAL 2.0, *supra* note 7, at 565.

22. Oona A. Hathaway, Rebecca Crootof, William Perdue, and Philip Levitz, *The Law of Cyber Attack*, 100 CALIFORNIA LAW REVIEW 817, 858 (2012); Geiß & Lahmann, *supra* note 7, at 632–33. For using belligerent reprisals to respond to cyberattacks in international armed conflicts, see SOLIS, *supra* note 7, at 692–95.

23. Michael N. Schmitt, *In Defense of Due Diligence in Cyberspace*, 125 YALE LAW JOURNAL FORUM 68, 78 (June 22, 2015), https://www.yalelawjournal.org/forum/in-defense-of-due-diligence-in-cyberspace. On necessity, see TALLINN MANUAL 2.0, *supra* note 7, at 135–42.

24. Gabčikovo-Nagymaros Project (Hung. v. Slovk.), Judgment, 1997 I.C.J. Rep. 7, ¶ 51 (Sept. 25).

A. Distinction between Causality and Attribution

Causality is a purely factual issue and asks whether a cyber operation can be traced to a particular computer system and, even more importantly, to a particular person, group of persons, or entity.

In contrast, attribution of the conduct of a particular person, group of persons, or entity to a particular State is, primarily, a normative issue.[25] The legal rules of attribution serve the purpose of limiting the international responsibility of States. A State within the territory of which a person, group of persons, or entity carried out a certain activity can be held internationally responsible only if there is a sufficiently close link between the person, group of persons, or entity and the State. Whether a link is sufficiently close is also a factual issue, but it is also a judgment. For making this judgment, several normative criteria have been developed and, finally, laid down in the 2001 ILC Draft Articles on Responsibility of States. These criteria define which relationships between a person, group of persons, or entity are sufficiently close to the State to consider their conduct that of the State.

The relevant normative criteria include instances in which the person, group, or entity is a State organ,[26] and where the person, group, or entity has been empowered with governmental authority by the State.[27] Moreover, the conduct of a person or group of persons acting on the "instructions" or under the "specific directions"[28] or "effective control"[29] of the State is an act of the State.[30] Whether these normative criteria are met depends, of course, on certain facts, which, if contested, must be proved.

In particular, efforts to attribute private conduct to States through the effective control test[31] face difficult factual challenges since in the

25. ILC Draft Articles with Commentary, *supra* note 3, at 38–39.

26. *Id.* art. 4.

27. *See id.* art. 5.

28. *Id.* at 48.

29. Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Judgment, 1986 I.C.J. Rep. 14, ¶ 115 (June 27) [hereinafter Nicaragua]; ILC Draft Articles with Commentary, *supra* note 3, at 47–48.

30. *See* ILC Draft Articles with Commentary, *supra* note 3, art. 8.

31. *Compare* ILC Draft Articles with Commentary, *supra* note 3, art. 8 ("The conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct."), *with* Nicaragua, *supra* note 29, ¶ 115

cyber context a party must demonstrate that the State exercised effective control over the "operations in the course of which the alleged violations were committed."[32] It may be similarly difficult to demonstrate that a private actor acted on the instructions of a State, a fact, which, if established, would, attribute the conduct to that State.[33] Under all three tests—effective control, specific directions, instructions—it has to be proven that "'effective control' was exercised, or that the State's instructions were given, for each operation in which the alleged violations occurred."[34] When applied to a cyber operation, the conduct complained of (e.g., launching the malware) must have been an "integral part" of the operation directed or controlled by, or carried out on the instructions of, the State.[35] It is not sufficient to show that control was exercised, or directions or instructions were given, "generally in respect of the overall actions taken by the persons or groups of persons having committed the violations."[36] Rather, the particular "physical acts constitutive of [the breach of an international obligation]" must have been "carried out wholly or in part, on the instructions or directions of the State, or under its effective control."[37]

B. Challenges to Causality and Attribution in Cyberspace

Due to the speed, anonymity, and ever-growing deceptive practices of cyber operations, the identification of the actor who launched a cyber operation poses a multitude of causality problems. Today, it is simple to hide one's tracks and traces of actions in the Internet and, as a further complication,

("For this conduct to give rise to legal responsibility of the United States, it would in principle have to be proved that that State had effective control of the military or paramilitary operations in the course of which the alleged violations were committed."); *see also* Application of the Convention on the Prevention and Punishment of the Crime of Genocide, *supra* note 3, ¶¶ 399–407.

32. Nicaragua, *supra* note 29, ¶ 115.

33. *See* ILC Draft Articles with Commentary, *supra* note 3, art. 8.

34. Application of the Convention on the Prevention and Punishment of the Crime of Genocide, *supra* note 3, ¶ 400.

35. *See* ILC Draft Articles with Commentary, *supra* note 3, at 47.

36. Application of the Convention on the Prevention and Punishment of the Crime of Genocide *supra* note 3, ¶ 400. The ICJ rejected the International Criminal Tribunal for the Former Yugoslavia's overall control test. *Id.* ¶¶ 402–06.

37. *Id.* ¶ 401.

conceal one's identity.[38] Anonymization tools like the TOR-browser[39] are available to anyone who can access the Internet and their use has become very common, especially for criminals trading on the Dark Net.[40] And while digital forensics continue to improve, traceability remains difficult if the only source of information is solely technical data. Without additional information gathered through intelligence, it remains difficult to know what person, group of persons, or entity initiated the cyber operation. In short, it is still nearly impossible to trace a cyber operation with absolute certainty to a particular computer system and, more importantly, to its author.[41]

Similar problems arise with regard to the facts necessary to establish the factual prerequisites of attribution of private conduct to a State under all three tests.[42] As Banks notes, "state and non-state cyber threats now often blend and merge, as privateers operate as surrogates for states and provide cover for state-based actors."[43] In particular, States may use proxies, defined here as "non-state actors with comparatively loose ties to governments,"[44] to commit malicious cyber operations without acting under the effective control, specific direction, or instructions of the government.[45] For ex-

38. Geiß & Lahmann, *supra* note 7, at 623–27.

39. *See The Onion Router (Tor)*, TECHOPEDIA, https://www.techopedia.com/definition/4141/the-onion-router-tor (last visited Nov. 21, 2019).

40. EUROPEAN LAW ENFORCEMENT AGENCY, IOCTA 2016: INTERNET ORGANISED CRIME THREAT ASSESSMENT 47–48 (2016).

41. Constantine Antonopoulos, *State Responsibility in Cyberspace*, *in* RESEARCH HANDBOOK ON INTERNATIONAL LAW AND CYBERSPACE, *supra* note 1, at 55, 62; Geiß & Lahmann, *supra* note 7, at 623–27; Martin Ney & Andreas Zimmermann, *Cyber-Security Beyond the Military Perspective: International Law, 'Cyberspace' and the Concept of Due Diligence*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 51, 56 (2015); WISSENSCHAFTLICHE DIENSTE DES BUNDESTAGS, WD 2–3000–038/15, ANWENDBARKEIT DES HUMANITÄREN VÖLKERRECHTS AUF COMPUTERNETZWERKOPERATIONEN UND DIGITALE KRIEGSFÜHRUNG (CYBER WARFARE) [SCIENTIFIC SERVICES OF THE BUNDESTAG, 2–3000–038/15, APPLICABILITY OF INTERNATIONAL HUMANITARIAN LAW TO COMPUTER NETWORK OPERATIONS AND DIGITAL WARFARE (CYBER WARFARE)] 10–11 (2015) [hereinafter CYBER WARFARE]. For a more comprehensive analysis of traceability problems, see SCHULZE, *supra* note 4, at 38–47.

42. *See supra* notes 31–37 and accompanying text.

43. Banks, *supra* note 7, at 274.

44. TIM MAURER, CYBER MERCENARIES: THE STATE, HACKERS, AND POWER 5 (2017).

45. *See supra* notes 26–30 and accompanying text.

ample, news media often report Russian hackers conduct cyber operations, but then stop short of claiming that the group acted under direct orders of the Russian government.[46]

C. Burden of Proof

The production of evidence establishing causality and attribution is the responsibility of the State that bears the burden of proof. When addressing a contested fact, the burden falls on the State that would benefit from the fact were it to be proved. Whether a particular fact is beneficial to a State depends on the legal claim raised by the State. If the claim asserted by the State only survives if a particular contentious fact is true, the burden of proof regarding this fact rests with the State asserting the claim. This placement of the burden aligns with the maxim *affirmanti incumbit probatio* because a State typically will assert only those facts that support its claim.[47] Indeed, the ICJ has held that "[a]s a general rule it is for the party which alleges a fact in support of its claims to prove the existence of that fact."[48]

Under this rule, if a State claims to have used force in the exercise of the right of self-defense, it must prove that it suffered an armed attack.[49]

---

46. One of the groups attributed to Russian intelligence services is APT 28, more popularly known as "fancy bear." Fabian A. Scherschel, *Bundestags-Hack: Angriff mit gängigen Methoden und Open-Source-Tools* [*Bundestag-Hack: Attack with Common Methods and Open-Source Tools*], HEISE ONLINE (Mar. 7, 2016), https://www.heise.de/security/meldung/Bundestags-Hack-Angriff-mit-gaengigen-Methoden-und-Open-Source-Tools-3129862.html. For a report detailing activity directed by the Russian government, see Press Release, Director of National Intelligence, Joint DHS, ODNI, FBI Statement on Russian Malicious Cyber Activity (Dec. 29, 2016), https://www.dni.gov/index.php/newsroom/press-releases/press-releases-2016/item/1616-joint-dhs-odni-fbi-statement-on-russian-malicious-cyber-activity.

47. In contrast, the maxim *onus probandi incumbit actori* is not sufficiently precise to express the doctrinal essence of the burden of proof in cyberspace cases and should, therefore, play no role in determining which party bears the burden. *See* ROBERT KOLB, THE ELGAR COMPANION TO THE INTERNATIONAL COURT OF JUSTICE 235 (2014).

48. Ahmadou Sadio Diallo (Guinea v. Dem. Rep. Congo), Merits, Judgment, 2010 I.C.J. Rep. 639, ¶ 54 (Nov. 30); *see also* Ahmadou Sadio Diallo (Guinea v. Dem. Rep. Congo), Compensation, Judgment, 2012 I.C.J. Rep. 324, ¶ 15 (June 19).

49. *See, e.g.*, Oil Platforms (Iran v. U.S.), Judgment, 2003 I.C.J. Rep. 161, ¶¶ 51, 57 (Nov. 6); *see also* CYBER WARFARE, *supra* note 41, at 12. *Cf.* U.N. Charter art. 51(1) ("Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security.").

It must also prove that the attack is attributable to the State against which force was used, that is, the State of origin of the attack.[50]

Thus, in the cyber context, to attribute adverse cyber operations to a State, the injured State must establish the following facts: (1) the identity of the person, group of persons, or entity authoring the operation; and (2) that the author or authors' conduct is attributable to that State.[51] The burden of proof is discharged only if the victim-State can demonstrate a causal link between the cyber operation, its adverse effects, and its author or authors.[52] In addition, if the author is a private individual or non-State actor, the injured State must prove that the operation was effectively controlled by, or carried out under the specific directions or instructions of the State against which it acted.

The burden of proof may be different, however, in instances where the response is a countermeasure against the State of origin of the adverse cyber operation. Depending on the circumstances of the individual case, the injured State may have to prove only that the State of origin breached its duty to prevent significant transboundary harm by not taking reasonable measures to prevent imminent, or stop ongoing, adverse cyber operations carried out from within its territory.[53]

Assuming the victim-State has produced evidence on another State's responsibility for a cyber operation, the question then arises whether the evidence is sufficient to discharge the burden of proof. This question has to be decided in light of the applicable standard of proof.

---

50. Oil Platforms, *supra* note 49, ¶¶ 51, 57, 59, 61, 71–72; *see also* Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion, 2004 I.C.J. Rep. 136, ¶ 139 (July 9); Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), Judgment, 2005 I.C.J. Rep. 168, ¶ 146 (Dec. 19); Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 239. For a highly critical view of the ICJ's jurisprudence, see SOLIS, *supra* note 7, at 686 ("Today, the court's view is essentially disregarded.").

51. For a comprehensive analysis concerning the methods of proof, that is, the kinds of evidence a party may produce to discharge its burden of proof in the context of cyber operations, see Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 254–64.

52. *Id.* at 243; CYBER WARFARE, *supra* note 41, at 12.

53. Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 239.

D. Standard of Proof

Green defines the standard of proof as "the quantum of evidence necessary to substantiate the factual claims by the parties."[54] Meeting this standard requires a party to persuade a court or trier of fact that its claims are true.[55] Of course, truth cannot be established objectively. Rather, the standard of proof denotes the degree of probability that must be achieved for the trier of facts to determine the factual allegation is correct.[56] Whether the required degree of probability and, therefore, the standard of proof is met depends on a deliberative assessment by the trier of fact based on the evidence submitted.[57]

1. Standard of Proof on the National Level

On the national level, the standards of "clear and convincing" and "beyond a reasonable doubt" are well-established concepts, with "beyond a reasonable doubt" being the highest standard of proof.[58] The lowest standard is "prima facie evidence," which requires merely indicative or plausible proof.[59] An intermediate standard is "preponderance of probability," under which the existence of the fact to be proved must be more likely than not, or, in other words, that it is reasonably probable.[60] Whereas common law jurisdictions apply all three standards of proof depending on the situation, civil law jurisdictions tend to limit the standard of proof to a single general rule: the judge must be "convinced" or "fully convinced" that a disputed fact is true.[61]

---

54. James A. Green, *Fluctuating Evidentiary Standards for Self-Defence in the International Court of Justice*, 58 INTERNATIONAL AND COMPARATIVE LAW QUARTERLY 163, 165 (2009).

55. *See* MARKUS BENZING, DAS BEWEISRECHT VOR INTERNATIONALEN GERICHTEN UND SCHIEDSGERICHTEN IN ZWISCHENSTAATLICHEN STREITIGKEITEN [THE LAW OF EVIDENCE BEFORE INTERNATIONAL COURTS AND ARBITRAL TRIBUNALS IN INTER-STATE DISPUTES] 506 (2010).

56. *See id.* at 506–07.

57. *See. id.* at 510.

58. *Id.* at 507; Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 248.

59. Geiß & Lahmann, *supra* note 7, at 624; Green, *supra* note 54, at 166; *see also* P. CHANDRASEKHARA RAO & PHILIPPE GAUTIER, THE INTERNATIONAL TRIBUNAL FOR THE LAW OF THE SEA: LAW, PRACTICE AND PROCEDURE ¶ 4.141 (2018).

60. *See* Green, *supra* note 54, at 167.

61. Roscini's statement that "in civil law systems there are no specific standards of proof that judges have to apply because they are authorized to evaluate the evidence produced according to their personal convictions on a case-by-case basis" is inaccurate. Roscini,

The degree of probability required to satisfy the convinced or fully convinced standard may be further defined as a "probability excluding reasonable doubt" or "probability close to certainty."[62] However, there are exceptions where the judge may lower the standard of proof by procedural rules or substantive provisions of the applicable law.[63]

2. Standard of Proof on the International Level

International criminal courts strictly adhere to the proof beyond a reasonable doubt standard.[64] Other international courts, tribunals, and dispute settlement bodies fluctuate between different standards of proof, often without providing precise definitions of each standard or their scope of application.[65]

Since our analysis is restricted to the exercise of the right of self-defense and the use of countermeasures—fundamental aspects of public interna-

---

Evidentiary Issues in International Disputes, *supra* note 7, at 248. The "principle of free assessment of evidence" has to be distinguished from the standard of proof. For a discussion of the distinction between the two, see KOLB, *supra* note 47, at 234, 251–52. In Nicaragua, the ICJ followed the principle of free assessment of evidence. *See* Nicaragua, *supra* note 29, ¶ 60 ("[The Court] has freedom in estimating the value of the various elements of evidence."). The principle of free assessment of evidence means that the assessment of evidence is not governed by formal rules on how to carry out the assessment and, in particular, what weight has to be given to each kind of evidence. *See, e.g.*, KOLB, *supra* note 47, at 234. Consequently, the interrelationship between the burden of proof, the standard of proof, and the principle of free assessment of evidence is as follows: whether the burden of proof is discharged due to sufficient evidence substantiating, in light of the applicable standard of proof, a contested fact has to be decided by an estimation of the value, reliability, and weight of the evidence in accordance with the principle of free assessment of evidence.

62. BENZING, *supra* note 55, at 507–08; RAO & GAUTIER, *supra* note 59, ¶ 4.140.

63. For a German perspective, see, for example, Wolfgang Hau, *Europarechtliche Vorgaben zum Beweismaß im Zivilprozess* [*European Law Requirements on the Standard of Proof in Civil Proceedings*], *in* DOGMATIK IM DIENST VON GERECHTIGKEIT, RECHTSSICHERHEIT UND RECHTSENTWICKLUNG: FESTSCHRIFT FÜR HANNS PRÜTTING [DOGMATICS IN THE SERVICE OF JUSTICE, LEGAL CERTAINTY AND LEGAL DEVELOPMENT: FESTSCHRIFT FOR HANNS PRÜTTING] 325, 326–27 (Moritz Brinkmann et al. eds., 2018).

64. Colleen M. Rohan, *Reasonable Doubt Standard of Proof in International Criminal Trials*, *in* PRINCIPLES OF EVIDENCE IN INTERNATIONAL CRIMINAL JUSTICE 650, 650 (Karim A.A. Khan et al. eds., 2010).

65. *See, e.g.*, BENZING, *supra* note 55, at 515–16, 526, 543–44, 548–49; Green, *supra* note 54, at 165–66.

tional law—the most authoritative source among international judicial bodies with regard to standards of proof is the ICJ.[66] That Court seems to insist on as much judicial latitude as possible. As a former judge and president of the ICJ states, "The Court's prime objective as to standard of proof appears to have been to retain a freedom in evaluating the evidence, relying on the facts and circumstances of each case."[67] Indeed, the ICJ has not articulated even a general standard of proof to be applied in cases brought before it.[68] One plausible explanation for this lack of an articulated standard is that a truly international bench of judges, such as the ICJ, includes lawyers from various legal traditions, including both common law and civil law jurisdictions.[69] This inclusivity, in turn, may contribute to a reluctance to make express reference to national concepts of standard of proof, each of which has its own legal history, doctrinal structure, and normative context. In fact, the ICJ restricts itself to drawing "inspiration from both the Anglo-Saxon legal tradition and continental systems of civil law."[70]

Several studies of the ICJ's jurisprudence suggest that in cases in which the international responsibility of a party is at stake the standard of proof depends on the nature of the international obligation allegedly breached by that party.[71] If the obligation breached is of "exceptional gravity," such as a claim alleging a violation of the prohibition of genocide, the standard of proof is that the "evidence . . . is fully conclusive" and the Court has to be "fully convinced that [the] allegations . . . have been clearly

66. For a much broader analysis taking into account other international courts, tribunals, and dispute settlement mechanisms, see BENZING, *supra* note 55, at 512.

67. Rosalyn Higgins, President of the International Court of Justice, Speech to the Sixth Committee of the General Assembly 4 (Nov. 2, 2007), https://www.icj-cij.org/files/press-releases/3/14123.pdf; *see also* Peter Tomka & Vincent-Joël Proulx, *The Evidentiary Practice of the World Court*, *in* LIBER AMICORUM: IN HONOUR OF A MODERN RENAISSANCE MAN HIS EXCELLENCY GUÐMUNDUR EIRÍKSSON 361, 363 (Juan Carlos Sainz Borgo ed., 2017) (stating "the rule of thumb for evidentiary matters before the Court is flexibility") (noting that Judge Tomka is a current member of the ICJ and a former president of the Court).

68. *See, e.g.*, Green, *supra* note 54, at 166; KOLB, *supra* note 47, at 251; Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 248–49.

69. With regard to the ICJ, see Higgins, *supra* note 67, at 5. For a similar assessment of the International Tribunal for the Law of the Sea, see RAO & GAUTIER, *supra* note 59, ¶ 4.140 (noting that Rao was a member of the Tribunal from 1996 to 2017, while Gautier was registrar of the Tribunal from 2001 to 2019).

70. Tomka & Proulx, *supra* note 67, at 364.

71. *See* Green, *supra* note 54, at 167–68, 170; KOLB, *supra* note 47, at 251–52; Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 248–49.

established."[72] Moreover, the Court stated, "[t]he same standard applies to the proof of attribution for such acts."[73] Some scholars have suggested that the Court is satisfied with a lower standard of proof[74] when the claim is that a State "has breached its undertakings to prevent genocide and to punish and extradite persons charged with genocide"[75] in requiring "proof at a high level of certainty appropriate to the seriousness of the allegation."[76] This argument is not persuasive.[77] Still, "appropriate to the seriousness of the allegation" supports the view that the standard of proof applied by the ICJ varies with the seriousness of the alleged breach of international law.[78]

Our doctrinal suggestion, therefore, is twofold. First, the pertinent standard of proof is derived from a sliding scale.[79] The highest standard, requiring "fully conclusive evidence," would apply to breaches of exceptional gravity. Continuing down the scale, the next standard, "conclusive" or "convincing evidence," would apply when the charges are of

72. Application of the Convention on the Prevention and Punishment of the Crime of Genocide, *supra* note 3, ¶ 209.

73. *Id.*

74. *See, e.g.*, BENZING, *supra* note 55, at 517.

75. Application of the Convention on the Prevention and Punishment of the Crime of Genocide, *supra* note 3, ¶ 210.

76. *Id.*

77. In the authors' view, the charge of having committed genocide and the charge of not having prevented genocide are of comparable exceptional gravity. This conclusion can be derived from the ICJ's *Genocide* judgment. *See id.* ¶¶ 209–10 (noting that the Court referred to "the standard of proof appropriate to charges of exceptional gravity" without distinguishing between "the crime of genocide" and the obligation "to prevent genocide and to punish and extradite persons charged with genocide").

78. *See, e.g.*, BENZING, *supra* note 55, at 549–50; *cf.* Corfu Channel (U.K. v. Alb.), Judgment, 1949 I.C.J. Rep. 4, 17 (Apr. 9) ("A charge of such exceptional gravity against a State would require a degree of certainty that has not been reached here."). *But see* Tomka & Proulx, *supra* note 67, at 376 (concluding that "the usual standard of proof tends to align with 'proof by a preponderance of the evidence'"). However, they do not elaborate further on the standard of proof and do not refer to any judgment of the ICJ. The more likely standard is clear and convincing evidence. *See* BENZING, *supra* note 55, at 514–15 (suggesting that the regular standard of proof applied by the ICJ seems to be the clear and convincing evidence test). For a more nuanced approach, see Mary Ellen O'Connell, *Rules of Evidence for the Use of Force in International Law's New Era*, 100 AMERICAN SOCIETY OF INTERNATIONAL LAW PROCEEDINGS 44, 44–45 (2006).

79. Concerning the proposal of an evidentiary sliding scale, see also Green, *supra* note 54, at 166–68 and TALLINN MANUAL 2.0, *supra* note 7, at 82.

severe gravity. A "preponderance of the evidence" would apply in instances of medium gravity. Finally, "*prima facie* evidence" would apply in cases of minor gravity.[80] Second, the standard of proof applied as a function of the seriousness of the alleged breach of international law would also apply with regard to attribution, and, thus, also to the causality of the conduct allegedly amounting to a breach of international law.

This sliding scale of standards of proof would also apply to a State's claim that it was lawfully permitted to deviate from a norm of public international law.[81] In the context of adverse cyber operations, possible responses include claims to take countermeasures or to exercise the right of self-defense.[82] Countermeasures are legally permissible responses to most breaches of international obligations[83] and may deviate from almost any international obligation.[84] Measures of self-defense are legally permissible responses to an armed attack, and may deviate from the prohibition of the use of force.[85] The application of the sliding scale of standards of proof proposed above depends on the gravity, or seriousness, of the deviation from public international law. As a result, the standard of proof is typically high in cases of self-defense and lower in cases of countermeasures. This means that the victim-State acting in self-defense through use of force (thus deviating from the prohibition of use of force) has to meet, as a rule, a higher standard of proof as regards its claim that it is, or has been, the victim of an armed attack by another State. Whereas the victim-State taking countermeasures below the threshold of use of force (thus deviating from norms *other than* the prohibition of use of force) has to meet a lower standard of proof

80. Such a flexible approach to the applicable standard of proof, taking into account the gravity of the breach of the international obligation, seems justified because the evidence is intertwined with the claim and, therefore, with the substantive rules on which the claim is based. Further, the very purpose of the standard of proof is to determine whether a disputed fact can be considered true. This determination includes an assessment that should take into account the legal consequences if the fact is held to be true. *See* BENZING, *supra* note 55, at 549–50.

81. *See also id.* at 520.

82. Dinstein, *supra* note 7, at 100, 102 (characterizing this exercise as "forcible countermeasures").

83. *See* ILC Draft Articles with Commentary, *supra* note 3, art. 22.

84. *See id.* art. 50(1)(a) (noting that the prohibition on the use of force remains in effect for countermeasures).

85. *See id.* art. 21 ("The wrongfulness of an act of a State is precluded if the act constitutes a lawful measure of self-defence taken in conformity with the Charter of the United Nations.").

as regards its claim that another State has breached an international obligation owed to the victim-State.[86]

E. Standard of Proof for Claims of Self-Defense

1. Generally Applicable Standard of Proof

Under the above analysis, determination of the standard of proof concerning claims of acting, or having acted, in self-defense as an exception to the general prohibition on the use of force[87] must take into account that the prohibition is widely accepted as a *jus cogens* rule.[88] Hence, a deviation from the prohibition of the use of force would be a case of exceptional gravity warranting the highest standard of proof (fully convinced).[89] Accordingly, a victim-State engaging in self-defense must have fully conclusive evidence establishing that an armed attack has occurred and that the attack is attributable to the attacking State. However, with regard to the prohibition of the use of force, one must "distinguish the most grave forms of the use of force (those constituting an armed attack) from other less grave forms."[90] Therefore, the highest standard of proof should apply only to claims of self-defense if the measures taken in self-defense amount to the most grave forms of the use of force. In case of measures taken by the victim-State against the attacker-State that are less grave forms of the use of force, conclusive or convincing evidence would be the appropriate standard of proof. Certainly, a State that is the victim of an armed attack, which responds with measures well below the threshold of the most grave forms of the use of force, should not be

86. *See id.* art. 49(1).

87. Nicaragua, *supra* note 29, ¶ 50.

88. *See especially* ILC Draft Articles with Commentary, *supra* note 3, at 247. To date, the ICJ has not expressly embraced the classification of the prohibition of the use of force as a *jus cogens* norm. *But see* Nicaragua, *supra* note 29, ¶ 190 (noting that when it examined the prohibition of the use of force under customary international law, the Court referred to the ILC, which had stated that the prohibition of the use of force had become part of the corpus of *jus cogens*).

89. *Cf.* Green, *supra* note 54, at 169 ("The very nature of the use of military force heightens the need for strict evidentiary requirements in respect of legal claims justifying such actions."); Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 250 ("As self-defense is an exception to the prohibition of the use of force, in particular, the standard of proof should be high enough to limit its invocation to exceptional circumstances and thus avoid abuses.").

90. Nicaragua, *supra* note 29, ¶ 191.

expected to meet the highest standard of proof in defending the lawfulness of its action.[91]

Admittedly, the ICJ seems to apply an intermediate standard of proof in self-defense cases without distinction as to the gravity of the amount of force used by the victim-State.[92] While expressly rejecting a "balance of evidence" test,[93] in *Oil Platforms*, the Court found the evidence presented by the United States "inconclusive" regarding Iran's responsibility for the naval mine struck by a U.S. warship or the missile attack on a U.S.-reflagged merchant vessel.[94] In that same judgment, the Court indicated conclusive evidence constituted evidence that was more than merely "suggestive" or "highly suggestive."[95] This finding is consistent with earlier statements of the Court indicating convincing evidence is the standard of proof to be applied in *jus ad bellum* cases involving actions taken by military forces.[96] Thus, as noted earlier,[97] one should hesitate to equate this international standard of proof (convincing evidence) with the national standard of clear and convincing evidence found in common law jurisdictions.[98]

2. Standard of Proof for Cyber Operations

Hence, based on the aforementioned ICJ jurisprudence, if the victim-State of an adverse cyber operation responds with the use of force against the State of origin, it must collect, and if necessary present sufficient evidence to meet the Court's convincing evidence standard of proof.[99] Thus, the evidence must convincingly establish the victim-State was, or continues to be,

---

91. This reasoning aligns with the sliding scale test, *see supra* notes 81–84 and accompanying text.

92. Green, *supra* note 54, at 173 ("A strict 'beyond a reasonable doubt' standard seems too onerous when it is considered that States making a genuine claim of self-defence will be faced with a *defensive necessity* for a military response.").

93. Oil Platforms, *supra* note 49, ¶ 57.

94. *Id.* ¶¶ 71–72.

95. *Id.* ¶¶ 59, 71.

96. *See* Corfu Channel, *supra* note 78, at 16–17; Nicaragua, *supra* note 29, ¶ 29; Armed Activities on the Territory of the Congo, *supra* note 50, ¶ 72.

97. *See supra* notes 69–70 and accompanying text.

98. *See* O'Connell, *supra* note 78, at 45; Green, *supra* note 54, at 172–74; Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 249–50.

99. *See, e.g.*, Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 252.

the object of an adverse cyber operation amounting to an armed attack that is attributable to the State of origin.

Of course, in some instances, it may be impossible for a State injured by an adverse cyber operation to discharge its burden of proof concerning causality and attribution.[100] Nevertheless, the technical and intelligence gathering complexities inherent in establishing causality and attribution in these cases do not warrant a lowering of the standard of proof generally applicable in *jus ad bellum* cases. Indeed, there is no persuasive argument for applying different standards of proof based on the manner in which the injury was inflicted. In all cases, a deviation from the prohibition of the use of force based on a claim of lawful self-defense must be assessed under the same standard of proof.[101] As an example, in *Oil Platforms*, the ICJ required the United States—the victim-State of an armed attack—to present conclusive evidence of Iranian culpability despite the challenges for doing so.[102] This same standard should apply for States responding to adverse cyber operations.

F. Standard of Proof for Cyber Countermeasures

1. Generally Applicable Standard of Proof

If the adverse cyber operation is not an armed attack within the meaning of Article 51 of the U.N. Charter, then there is no right to engage in self-defense. In such instances, the taking of non-forcible countermeasures,[103] as defined by Article 22 of the Draft Articles, may be the appropriate response.[104] However, countermeasures are not available to the injured State if

100. *See supra* notes 38–46 and accompanying text.

101. *See* Green, *supra* note 54, at 169 (arguing "a consistent evidentiary standard is desirable for self-defence"); *see also* Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 249.

102. *See* Oil Platforms, *supra* note 49, ¶¶ 50, 63, 67, 72.

103. In contrast to self-defense measures, these actions may be characterized as "forcible countermeasures." *See* Dinstein, *supra* note 7, at 100, 102.

104. However, it is important to remember that the sole purpose of a countermeasure must be "the restoration of a condition of legality as between the injured State and the responsible State," that is, countermeasures must be "taken as a form of inducement, not punishment." ILC Draft Articles with Commentary, *supra* note 3, at 130–31; *see also* TALLINN MANUAL 2.0, *supra* note 7, at 112, 116–22.

the cyber operation cannot be attributed to the State of origin. In that situation, the victim-State is restricted to protests to that State for its failure to comply with its duty to prevent significant transboundary harm.[105]

An injured State must prove each fact necessary to establish that the legal requirements for taking countermeasure are met.[106] In the *Gabčikovo-Nagymaros* judgment, the ICJ required a State taking countermeasures to present evidence demonstrating that the offending State "is responsible for an internationally wrongful act."[107] This condition implies that the internationally wrongful act must have already occurred,[108] and that it is attributable to the other State under one of the circumstances set forth in Articles 4 through 10 of the Draft Articles.[109] Accordingly, as in the case of self-

105. *See* TALLINN MANUAL 2.0, *supra* note 7, at 113.

106. For an extensive discussion of the permissibility of countermeasures in response to adverse cyber operations, see *id.* at 111–42.

107. Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 83; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 49(1).

108. Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 83.

109. The additional prerequisites of a lawful countermeasure are that the countermeasure must be a response to the previously committed internationally wrongful act. *Id.* Therefore, the countermeasure must be "directed against that State" that is internationally responsible for the previous internationally wrongful act. *Id.*; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 49(1). In addition, the purpose of the countermeasure "must be to induce the wrongdoing State to comply with its obligations under international law." Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 57; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 49(1). In other words, the countermeasure's aim must be "to procure cessation and reparation." *Id.* at 128. Accordingly, countermeasures "are essentially temporary measures, taken to achieve a specified end, whose justification terminates once the end is achieved." *Id.* at 129. Hence, as has been held by the ICJ, any countermeasure "must therefore be reversible." Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 87. This requirement of reversibility casts doubt on whether hack backs can always be considered lawful countermeasures. With regard to the problem of reversibility, see TALLINN MANUAL 2.0, *supra* note 7, at 119. Moreover, "the injured State must have called upon the State committing the wrongful act to discontinue its wrongful conduct or to make reparation for it." Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 84; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 52(1). Finally, "the effects of a countermeasure must be commensurate with the injury suffered, taking account of the rights in question." Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 85; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 51. Therefore, the lawfulness of a countermeasure depends on its being "proportionate." Gabčikovo-Nagymaros Project, *supra* note 24, ¶ 87; *see also* ILC Draft Articles with Commentary, *supra* note 3, art. 51.

defense, the State taking countermeasures must prove the facts substantiating causality, as well as the factual basis for attributing the wrongful act to the State against which the countermeasures are directed.

As discussed above, the pertinent standard of proof depends on the gravity of the breach of the international obligation that the injured State owes to the other State.[110] Countermeasures are "measures that would otherwise be contrary to the international obligations of an injured State vis-à-vis the responsible State, if they were not taken by the former in response to an internationally wrongful act by the latter in order to procure cessation and reparation."[111] Accordingly, it is not the gravity of the previously committed breach of the international obligation by the responsible State that determines the standard of proof the injured State must meet to justify its countermeasure.

The gravity of the injured State's departure from the international obligations it owes to the responsible State depends on the nature of the international obligations and the extent of their impairment through the taking of a countermeasure. All international obligations of the injured State vis-à-vis the responsible State may be breached by a countermeasure except those, which are listed in Article 50 of the Draft Articles.[112] Among these sacrosanct international obligations are the prohibition on the threat or use of force and obligations under norms of *jus cogens*.[113] Therefore, the international obligations from which the injured State intends to derogate by the taking of countermeasures will be those of less fundamental importance. Hence, the standard of proof concerning factual allegations by the injured State in support of its claim to have taken countermeasures is not typically the highest standard of proof of fully conclusive evidence, but conclusive or convincing evidence, or a preponderance of evidence. The choice between these two standards of proof depends not only on the nature of the obligation impaired by the countermeasure, but also on the magnitude of the violation of the international obligation.

110. *See supra* notes 71–86 and accompanying text.

111. ILC Draft Articles with Commentary, *supra* note 3, at 131.

112. *Id.* art. 50.

113. *Id.*; *see also* TALLINN MANUAL 2.0, *supra* note 7, at 122–26.

2. Standard of Proof for Cyber Operations

The most effective countermeasures to adverse cyber operations, and the most intensely debated, are "active defenses," also known as "hack-backs."[114] Examples include "programs that send destructive viruses back to the perpetrator's machine or packet-flood the intruder's machine."[115] An in-depth discussion of the legality of active defenses is beyond the scope of this article. Here, it is assumed that, while their use must be analyzed in each individual case, active defenses are in general justifiable countermeasures.

Like with all countermeasures, the standard of proof when active defense measures are taken depends on the nature of the international obligation impaired. And, as with other countermeasures, active defenses may not violate the prohibition on the threat or use of force. If a countermeasure in the form of an active defense were an act of the use of force, it would trigger the standard of proof to be applied generally in *jus ad bellum* cases, that is, conclusive evidence.[116] In this regard, Judge Simma, in his separate opinion in *Oil Platforms*,[117] observed that forcible "proportionate defensive measures" in response to hostile acts could be justified as falling "short of" an Article 51 "(full-scale) self-defence."[118] Some legal experts applied this reasoning to adverse cyber operations that do not amount to an armed attack arguing that forcible cyber countermeasures that do not cross the threshold of the most grave forms of the use of force are lawful.[119] But even applying Judge Simma's "lower-level, smaller-scale proportionate defensive measures"[120] criteria in response to adverse cyber operations might be considered a use of force, albeit a less grave form of use of force. And as a use of force, the

114. *See supra* notes 19–22 and accompanying text.

115. Sklerov, *supra* note 7, at 25.

116. *See supra* notes 87–98 and accompanying text.

117. Oil Platforms (Iran v. U.S.), Judgment, 2003 I.C.J. Rep. 161, 324, ¶¶ 51, 57 (Nov. 6) (separate opinion by Simma, J.).

118. *Id.* at 331–32, ¶ 12.

119. *See* TALLINN MANUAL 2.0, *supra* note 7, at 125; *see also* Benedikt Pirker, *Territorial Sovereignty and Integrity and the Challenges of Cyberspace*, *in* PEACETIME REGIME FOR STATE ACTIVITIES IN CYBERSPACE, *supra* note 7, at 189, 213; Sklerov, *supra* note 7, at 37.

120. Oil Platforms (Iran v. U.S.), Judgment, 2003 I.C.J. Rep. 161, 324, ¶ 12 (separate opinion by Simma, J.).

standard of proof to justify an active defense would certainly have to be that of conclusive evidence.[121]

Whether active defenses are below the use of force threshold is ultimately a question of degree depending on the circumstances of the individual case and particularly the factual consequences arising from their use.[122] If, due to their small scale and limited effects, they do not constitute a use of force, they may nevertheless contravene the principle of non-intervention. This does not mean that active defenses that are not a use of force are always an intervention. A wrongful intervention is defined by the "element of coercion," which "forms the very essence of [a] prohibited intervention."[123] The only legitimate objective of active defenses is to induce the State of origin to cease adverse cyber operations and return to compliance with the law.[124] Thus, active defenses typically have an element of coercion since they are in effect law enforcement measures.[125] More importantly, active defenses may damage the information technology (IT) systems at which they are directed. Accordingly, we suggest that the applicable standard of proof should generally be conclusive evidence.[126] The standard should be lowered to a preponderance

121. See the ICJ's jurisprudence in the *jus ad bellum* cases *supra* notes 92–96 and accompanying text.

122. For an extensive analysis of the threshold of the use of force with regard to cyber operations, see TALLINN MANUAL 2.0, *supra* note 7, at 330–37.

123. Nicaragua, *supra* note 29, ¶ 205.

124. ILC Draft Articles with Commentary, *supra* note 3, art. 49(1).

125. Admittedly, in its *Nicaragua* decision, the ICJ held that coercion resulting in a violation of the principle of non-intervention must be related to "choices" in "matters in which each State is permitted, by the principle of State sovereignty, to decide freely." Nicaragua, *supra* note 29, ¶ 205. In this regard, the Court refers to "the choice of a political, economic, social and cultural system, and the formulation of foreign policy." *Id.* Accordingly, one may argue that such choices are, typically, not the target of active defenses. *See* Dörr, *supra* note 10, at 90. On the other hand, the Court "define[d] only those aspects of the principle [of non-intervention] which appear[ed] to be relevant to the resolution of the dispute." Nicaragua, *supra* note 29, ¶ 205. Hence, the Court seemingly did not intend to interpret the principle of non-intervention exhaustively. Thus, other forms of coercion could constitute an intervention as well. For an extensive analysis of what might constitute "cyber intervention," see Terry D. Gill, *Non-Intervention in the Cyber Context*, *in* PEACETIME REGIME FOR STATE ACTIVITIES IN CYBERSPACE, *supra* note 7, at 217, 232–36; *see also* TALLINN MANUAL 2.0, *supra* note 7, at 312–27.

126. Even if such "active defences" were not considered an unlawful "intervention," they might constitute "a different form of violation of State sovereignty." *See* Dörr, *supra* note 10, at 90. For a discussion of violations of State sovereignty through

of the evidence if the victim-State's countermeasures have only a minor effect on IT systems and are designed to warn the State of origin of more severe actions to follow if the adverse cyber operations continue.[127]

G. Proof of Attribution

In many, if not most cases, the standard of proof will be convincing evidence, a rather high standard of proof. This standard does not rule out the consideration of indirect evidence.[128] Accordingly, depending on its reliability, value, and weight, indirect evidence may be sufficient to meet the convincing evidence standard.[129] In *Corfu Channel*, the ICJ correctly held that "indirect evidence is admitted in all systems of law, and its use is recognized by

---

cyber operations, see, for example, Wolff Heintschel von Heinegg, *Territorial Sovereignty and Neutrality in Cyberspace*, 89 INTERNATIONAL LAW STUDIES 123, 129 (2013); *see also* Michael N. Schmitt, *Cyber Activities and the Law of Countermeasures*, *in* PEACETIME REGIME FOR STATE ACTIVITIES IN CYBERSPACE, *supra* note 7, at 659, 685 (arguing that the standard of "reasonable certainty" should apply when determining "factual attribution").

127. Such countermeasures could be called "show of force countermeasures."

128. Indirect evidence or circumstantial evidence must not be confused with the applicable standard of proof. For an instance where this occurred, see O'Connell, *supra* note 78, at 45; *see also* Antonopoulos, *supra* note 41, at 63–64; Green, *supra* note 54, at 175; Pirker, *supra* note 119, at 206; Christian Walter, *Obligations of States Before, During and After a Cyber Security Incident*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 67, 82 (2015) [hereinafter Walter, *Obligations of States*]. Whether indirect evidence is admissible, and what value and weight it should receive, are questions regarding the assessment of evidence. *See* KOLB, *supra* note 47, at 243 ("[A]n inference is nothing more than a matter of understanding the evidence and its interpretation."). Of course, the assessment of indirect evidence in the absence of any direct evidence may lead a court to conclude that the party has not discharged its burden of proof. Accordingly, the acceptance of indirect evidence does not in itself mean that the applicable standard of proof was lowered. Rather, whether indirect evidence is sufficient to discharge the burden of proof has to be decided in light of the applicable standard of proof.

129. Admittedly, in *Corfu Channel* the ICJ held that "[t]he proof may be drawn from inferences of fact, provided that they leave *no room* for reasonable doubt." Corfu Channel, *supra* note 78, at 18. In our opinion, this passage does not imply that the Court opted for the highest standard of proof (beyond a reasonable doubt) if only circumstantial evidence is available. Rather, we suggest that the Court applied the highest standard of proof because of the serious allegation by the United Kingdom that Albania "knew that the . . . minefield was lying in . . . its territorial waters" and that no notice of the mines was provided to shipping vessels generally or British naval vessels specifically. *Id.* at 10.

international decisions."[130] The Court further explained that indirect evidence "must be regarded as of special weight when it is based on a series of facts linked together and leading logically to a single conclusion."[131] Indirect evidence is of the utmost importance if the conduct that breached an international obligation took place in territory under the exclusive control of the State being held accountable due to the difficulty of collecting evidence. The Court explicitly acknowledged that, "[b]y reason of this exclusive control, the other State, the victim of a breach of international law, is often unable to furnish direct proof of facts giving rise to responsibility."[132]

1. Situation I: Cyber Operation Traceable to a State IT System

The factual situation in the *Corfu Channel* case is analogous to the situation in many cases involving adverse cyber operations. Here, the victim-State may be able to use digital forensics and "inferences of fact and circumstantial evidence"[133] to trace the cyber operation to a State IT system located within a State-owned or State-controlled facility[134] of the State of origin.[135] The missing causal link is the conduct of the person, group of persons, or entity that launched the cyber operation.[136] Without knowledge of the specific conduct and the individuals or entity involved, it is impossible to apply the rules of attribution as reflected in Articles 4 to 10 of the Draft Articles.[137]

Nevertheless, when the operation has definitely been launched from a State IT system located within a State-owned or State-controlled facility, the victim-State has conclusively established that the adverse cyber operation was the responsibility of a State organ within the meaning of Article 4 of the Draft Articles. This conclusion is reasonably inferred from the facts that

130. *Id.* at 18.

131. *Id.*

132. *Id.*

133. *Id.*

134. One of the first in-depth studies on advanced persistent threats (APTs) tracked certain cyber operations to a building complex in China. MANDIANT, APT1: EXPOSING ONE OF CHINA'S CYBER ESPIONAGE UNITS (2013), https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf.

135. *See also* Walter, *Cyber Security*, *supra* note 14, at 690; Walter, *Obligations of States*, *supra* note 128, at 71–72. Walter bases his analysis on the distinction between cyber operations traced to governmental IT systems and cyber operations traced to private IT systems.

136. *See* Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 240.

137 *See* ILC Draft Articles with Commentary, *supra* note 3, arts. 4–10.

have been proven: only public officials, hence, persons having the status of a State organ or belonging to an entity having said status,[138] working in, or having access to the State facility could have initiated the cyber operation and did so in an official capacity.[139] In these circumstances, the injured State has discharged its burden of proof in establishing that it was authorized to respond with actions in self-defense and countermeasures.[140]

Admittedly, one may wonder whether it is possible to establish through mere inference that a cyber operation launched from a State IT system within a State-owned or State-controlled facility was conducted by a State organ. However, one has to take into account that it is, in accordance with the principle of free assessment of evidence, a deliberative judgment as to whether the applicable standard of proof has been met by circumstantial evidence. Such a deliberative judgment must consider the maxim *ultra posse nemo obligatur* (No one is obligated beyond what he is able to do). Hence, the victim-State cannot be required to submit direct evidence regarding the identity of the individuals operating the State IT system within the State-owned or State-controlled facility during the cyber operation, or evidence as to whether these individuals acted in the official capacity of State organs. It is practically impossible to produce such direct evidence. It is only possible for the State of origin to disclose which persons operated the relevant IT system when the cyber operation was launched.

Consequently, the burden of proof is shifted to the State of origin. That State may claim—and prove—that a State organ did not conduct the adverse

138. *Id.* art. 4(2).

139. *Id.* at 42. The same reasoning would apply if the adverse cyber operation could be traced to the premises of an entity endowed with elements of governmental authority within the meaning of Article 5 of the Draft Articles.

140. *See also* Pirker, *supra* note 119, at 205–06. Similar results, but through a different doctrinal avenue, may be achieved by using a rebuttable presumption. Here, if the adverse cyber operation can be traced to a State IT system, it is presumed that the cyber operation has been launched by a State organ and is therefore attributable to the State of origin. For such an approach, see Antonopoulos, *supra* note 41, at 62; Walter, *Cyber Security*, *supra* note 14, at 690; Walter, *Obligations of States*, *supra* note 128, at 72. *But see* von Heinegg, *supra* note 126, at 137 (concluding only knowledge of, but not attribution to, the State of origin can be presumed). Similarly, *Tallinn Manual 2.0* states, "the mere fact that a cyber operation has been launched or otherwise originates from governmental cyber infrastructure . . . is usually insufficient evidence for attributing the operation to that State." TALLINN MANUAL 2.0, *supra* note 7, at 91. However, the Manual also notes, "such usage can serve as an indication that the State in question may be associated with the operation." *Id.*

cyber operation.[141] For example, it may submit evidence establishing conclusively that a particular non-State actor (or a State organ of a third State) committed the adverse cyber operation through unlawful use, or hacking, of the State IT system. If the State of origin is unable to establish convincingly that a specific non-State actor introduced the malware into its IT system, it must then prove a so-called "negative fact," namely the absence of culpable conduct by its own State organs. In that regard, however, the legal principle *negativa non sunt probanda* or *factum negantis nulla probatio*, under which a party denying a fact is not bound to offer proof, is no longer a rule of international law.[142] Indeed, in *Nicaragua* the ICJ held that "[t]he evidence or material offered by Nicaragua in connection with the [United States'] allegation of arms supply [by Nicaragua to Salvadorians] has to be assessed bearing in mind the fact that, in responding to that allegation, Nicaragua has to prove a negative."[143] This passage, while not reviving the Roman rule, indicates that the evidentiary burden may be lowered when it comes to establishing a negative fact.[144]

Due diligence must also be considered. If the State of origin can prove convincingly that it took all reasonable measures to prevent the launching of adverse cyber operations from within State-owned or State-controlled facilities or the hacking of its IT system within those facilities, then it has discharged its burden of proof.[145] That is, it convincingly established the negative fact that the actor responsible for the adverse

---

141. For agreement, see Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 243 ("Once the burden has been discharged according to the appropriate standard, the burden shifts to the other litigant, who has to prove the contrary."); *see also* Pirker, *supra* note 119, at 206; Walter, *Cyber Security*, *supra* note 14, at 690; Walter, *Obligations of States*, *supra* note 128, at 72.

142. KOLB, *supra* note 47, at 244.

143. Nicaragua, *supra* note 29, ¶ 147.

144. Or, depending on the nature of the international obligation, when the disputing parties contest the performance of the disputed fact, the burden of proof may be reversed requiring the other party to prove the corresponding positive fact. For a case in which the ICJ addressed this approach, see Ahmadou Sadio Diallo, Merits, Judgment, *supra* note 48, ¶ 55 (concerning whether the Democratic Republic of the Congo, the respondent State, had complied with certain procedural guarantees).

145. This does not imply that the authors assume the existence of a general duty to prevent significant transboundary harm under public international law in the cyber context. Rather, the suggestion here is merely that a State may escape international responsibility for an adverse cyber operation traced to its State IT systems if that State has taken measures to prevent the use of its IT system for cyber operations harmful to

cyber operation was not, and could not have been, a State organ. The true actor had to have been a non-State actor or a third State not acting under the effective control, the specific directions, or the instructions of the State of origin.

Here, the injured State, having established the adverse cyber operation was launched from a State of origin's IT system located within a State-owned or State-controlled facility, and the State of origin having established that it had to have been authored by a non-State actor or third State, the issue becomes which State bears the burden of proof for a contentious non-proven fact. In this instance—a situation of *non liquet*—the facts to be proven are conduct (the launching of an adverse cyber operation) by a person, group of persons, or entity having the status of a State organ, or by non-State actors acting under the control, directions, or instructions of the State of origin.[146] The burden of proof for establishing these facts falls to the victim-State, which has failed in that regard since the State of origin has provided convincing evidence that, due to a lack of causation and attribution, it cannot be held responsible for the adverse cyber operation. Accordingly, the victim-State cannot respond through self-defense or take countermeasures.

2. Situation II: Cyber Operation Traceable to a Private IT System

If the victim-State convincingly establishes that the adverse cyber operation was launched from a private IT system within the territory of the State of origin, the analysis to determine State responsibility becomes much different. The inference that can be made when a State IT system within a State-owned or State-controlled facility is the source of the operation no longer applies. To the contrary, the obvious inference to make is that the adverse cyber operation resulted from private conduct.

---

other States. Concerning the contentious question of whether States are under a general duty to prevent significant transboundary harm arising from adverse cyber operations launched from within their territory. *Compare* Dörr, *supra* note 10, at 93–94 (affirming the existence of a duty), *with* TALLINN MANUAL 2.0, *supra* note 7, at 45–50 (finding no such duty). For a discussion of the difficulty of applying the due diligence principle in cyberspace, see Jutta Brunnée & Tamar Meshel, *Teaching an Old Law New Tricks: International Environmental Law Lessons for Cyberspace Governance*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 129, 140 (2015); *see also* Robert Kolb, *Reflections on Due Diligence Duties and Cyberspace*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 113, 120, 126 (2015).

146. ILC Draft Articles with Commentary, *supra* note 3, arts. 4, 8.

Under these facts, for purposes of attribution to the State of origin, the State injured by an adverse cyber operation must prove the conduct was carried out by a private person, group of persons, or entity under the effective control, specific direction, or the instructions of the State of origin within the meaning of Article 8 of the Draft Articles.[147] Alternatively, it would sustain its burden of proof if it could establish that, while a private IT system was utilized, the private IT system was operated by a State organ, as defined by Article 4 of the Draft Articles.[148] Again, it may be extremely difficult, if not impossible, for the victim-State to submit convincing evidence that these factual requirements for attribution are met. On the assumption that the victim-State has convincingly established that the cyber operation was initiated from a private IT system within the State of origin's territory, it still must prove additional facts to satisfy attribution. Namely, it must prove that certain conduct by specified actors being either State organs or non-State actors operating under the effective control, specific directions, or the instructions of the State of origin lie within the domain, or sphere, of the State of origin. Absent cooperation of that State, these facts are inaccessible to the injured State.[149] In such a situation, the victim-State is again limited to the production of indirect evidence, that is, proof of facts from which it can be inferred that the adverse cyber operation was conducted by either a State organ or by a non-State actor acting under the effective control, specific directions, or the instructions of the State of origin.[150]

One alternative would be to shift the burden of proof to the State of origin. Antonopoulos argues that if an

147. *See id.* art. 8. In the present article, we limit the analysis to private actors who are not members of terrorist-designated groups, such as Al-Qaida or ISIS. If the victim-State can convincingly establish that the adverse cyber operation was launched by terrorists or a terrorist group from within the State of origin's territory, the rules relating to terrorist attacks from abroad, as they have evolved following the 9/11 attacks, would apply.

148. *See* ILC Draft Articles with Commentary, *supra* note 3, art. 4. The State could also sustain its burden if it could establish that the adverse cyber operation came from a private IT system operated by an entity entrusted with governmental authority within the meaning of Article 5 of the ILC Draft Articles. *See id.* art. 5.

149. *Cf.* Corfu Channel, *supra* note 78, at 34–35 (noting that the Court explicitly rejected a right to intervention "by means of which the State intervening would secure possession of evidence in the territory of another State, in order to submit it to an international tribunal and thus facilitate its task").

150. The same argument would apply to a State-empowered entity within the meaning of Article 5 of the Draft Articles. *See supra* note 148 and accompanying text.

> injurious cyber activity can be traced to the territory of a single State . . . the best approach in the cyber context is to attribute hostile cyber acts to this State on the basis of a presumption of responsibility which may be rebutted by the State on the basis of evidence."[151]

However, there are a number of issues to consider in adopting this approach, and under no circumstances, should this approach apply to the exercise of self-defense or the taking of countermeasures.

First, the ICJ's jurisprudence does not support shifting the burden of proof in this manner.[152] In *Corfu Channel*, the ICJ explicitly held that "the mere fact of the control exercised by a State over its territory . . . by itself and apart from other circumstances . . . [does not] shift[] the burden of proof,"[153] even in cases not involving the use of force. Second, in case of a claim of self-defense, the burden of proof to present facts supporting the claim falls completely with the victim-State. This rule would also apply to a claim of the right to take countermeasures. Third, the use of force through the exercise of the right of self-defense cannot be premised on mere presumptions concerning causality and attribution. To do so could subject the State of origin to the use of force on little more than *prima facie* evidence. This is far too low a standard to allow breaches of the prohibition on the use of force. The same applies to the use of active defenses. Fourth, to rebut the presumption, the State of origin would be required to prove negative facts, namely that the private person, group of persons, or entity concerned was *not* acting under its effective control, specific directions, or its instructions, or that the private IT system was *not* operated by a State organ when the adverse cyber operation was initiated.[154]

---

151. Antonopoulos, *supra* note 41, at 64 (conceding that this approach "introduces a reversal of the burden of proof"). For rejection of such an approach, see SCHULZE, *supra* note 4, at 153; *see also* Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 246.

152. *See* Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 246–48; *cf.* KOLB, *supra* note 47, at 245–46 (discussing several approaches to meeting the standard of proof such as lowering the standard, softening or shifting the burden of proof, or allowing for a benevolent assessment of the indirect evidence of other international courts, tribunals, and quasi-judicial dispute settlement bodies).

153. Corfu Channel, *supra* note 78, at 18.

154. The same argument would apply to a State-empowered entity within the meaning of Article 5 of the Draft Articles. *See supra* note 148 and accompanying text.

Of course, both the burden of proof and the corresponding standard of proof must not "degenerat[e] into a *probatio diabolica* (a legal requirement to achieve an impossible proof)."[155] Accordingly, a State must not be placed in circumstances where its burden of proof is impossible, or almost impossible, to discharge against the applicable standard of proof. The purpose of procedural rules, as of any law, is to ensure justice.[156] Therefore, the interpretation and application of procedural rules (and thus, also evidentiary rules), should not prevent States from being held accountable for serious breaches of their international obligations. This applies in particular to those adverse cyber operations that are armed attacks within the meaning of Article 51 of the U.N. Charter, but also to other adverse cyber operations falling below this threshold.

One situation that could amount to a *probatio diabolica* is when an injured State convincingly establishes that the cyber operation can be traced to a private IT system located in the State of origin's territory, but it is unable to identify the actor whose conduct launched the adverse cyber operation with convincing evidence. Similar difficulties arise for establishing if the actor, when identified, was a State organ[157] or a private person, group of persons, or entity acting under the effective control, specific directions, or the instructions of, State of origin authorities.

To avoid a *probatio diabolica*, the State of origin is under a duty to cooperate in good faith with the victim-State.[158] As long as the State of origin cooperates in good faith, the victim-State will not have recourse to self-defense measures or active defenses. This duty of the State of origin to cooperate is an expression of its "obligation not to allow knowingly its territory to be used for acts contrary to the rights of other States."[159] Once the injured State has notified the State of origin that it was the victim of an adverse cyber operation emanating from a private

155. KOLB, *supra* note 47, at 245–46.

156. The very "idea" of law is justice. *See* GUSTAV RADBRUCH, RECHTSPHILOSOPHIE [LEGAL PHILOSOPHY] 30 (3d ed., 1932) ("Die Idee des Rechts kann nun keine andere sein als die Gerechtigkeit." ["The idea of law can now be none other than justice."] (translation by authors).

157. The same argument would apply to a State-empowered entity within the meaning of Article 5 of the Draft Articles. *See supra* note 148 and accompanying text.

158. Concerning the duty to cooperate, see Brunnée & Meshel, *supra* note 145, at 145; Dörr, *supra* note 10, at 97–98; Walter, *Cyber Security*, *supra* note 14, at 690; Walter, *Obligations of States*, *supra* note 128, at 81.

159. Corfu Channel, *supra* note 78, at 22.

IT system within the latter State's territory, the origin State has knowledge that its territory has been used for an act infringing upon the rights of other States.[160] This knowledge activates the State of origin's duty to prevent the continuation of such acts.[161] As the ICJ held in *Corfu Channel*, this duty includes a duty to cooperate.

> It is true that a State on whose territory or in whose waters an act contrary to international law has occurred may be called upon to give an explanation. It is also true that that State cannot evade such a request by limiting itself to a reply that it is ignorant of the circumstances of the act and of its authors. The State may, up to a certain point, be bound to supply particulars of the use made by it of the means of information and inquiry at its disposal.[162]

Accordingly, the injured State has to establish convincingly that it sought information from the State of origin "with sufficient specificity" on the operation of the private IT system at the time the adverse cyber operation began, and that the authorities of the State of origin "declined or failed to respond to such specific requests."[163]

If the State of origin is unwilling to cooperate or cooperates in bad faith, for example, by unduly delaying or subverting cooperation efforts or by proposing obscure conspiracy theories, the victim-State may infer from such behavior that the private IT system was either operated by State organs[164] or by private persons, groups of persons, or entities acting under the effective control, specific directions, or the instructions of the State.[165] This is not a presumption of causality and attribution, but an inference drawn from the

---

160. Kolb, *supra* note 145, at 123; *see also* von Heinegg, *supra* note 126, at 136; Schmitt, *In Defense of Due Diligence*, *supra* note 23, at 75–76, 79. Concerning knowledge as a constitutive element in the application of the due diligence principle, see TALLINN MANUAL 2.0, *supra* note 7, at 40–43.

161. *Tallinn Manual 2.0* derives the duty to stop from the due diligence principle. *See* TALLINN MANUAL 2.0, *supra* note 7, at 43–50; *see also* Ney & Zimmermann, *supra* note 41, at 64; Schmitt, *In Defense of Due Diligence*, *supra* note 23, at 79.

162. Corfu Channel, *supra* note 78, at 18.

163. Avena and Other Mexican Nationals (Mex. v. U.S.), Judgment, 2004 I.C.J. Rep. 12, ¶ 57 (Mar. 31).

164. The same argument would apply to a State-empowered entity within the meaning of Article 5 of the Draft Articles. *See supra* note 148 and accompanying text.

165. This outcome would not apply if the State convincingly established facts contrary to such an inference. *See* Roscini, *Evidentiary Issues in International Disputes*, *supra* note 7, at 268–69.

facts, therefore it relies on circumstantial evidence of causality and attribution. Such indirect evidence should, under these circumstances and in view of the rigorous and exhaustive efforts of the victim-State to discharge its burden of proof, be considered sufficient to meet the convincing evidence standard of proof. This conclusion aligns with the ICJ jurisprudence concerning evidence collection in difficult circumstances.[166] Accordingly, the victim-State would have to be considered to have discharged its burden of proof as regards causality and attribution of the adverse cyber operation to the State of origin. Hence, the victim-State could invoke its right to self-defense, or take low-level forcible countermeasures if all other prerequisites for a lawful exercise of these rights were met.[167]

If the State of origin cooperates in good faith, the injured State may not have recourse to either self-defense or cyber countermeasures. The full cooperation shows that the State adheres to its obligations under public international law. Hence, inducement to return to the rule of law through forcible countermeasures is no longer necessary, at least for the period of cooperation. More importantly, the injured State cannot be considered to have discharged its burden of proof if, despite the full cooperation of the State of origin, the attribution of the author of the adverse cyber operation to the State of origin cannot be shown. In this case, the victim-State has neither convincingly established that the adverse cyber operation was attributable to the State of origin nor convincingly established that the State of origin violated its "obligation not to allow knowingly its territory to be used for acts contrary to the rights of other States," as set forth in *Corfu Channel*.[168]

If, however, the adverse cyber operations resume, and these operations can be traced to the same private IT systems identified previously, the victim-State may infer that the State of origin knew, or should have known, that the private IT system had again been used to initiate adverse

166. Land, Island, and Maritime Frontier Dispute (El Sal. v. Hond., Nicar. intervening), Judgment, 1992 I.C.J. Rep. 351, ¶ 63 (Sept. 11).

167. Sklerov arrives at the same conclusion, but by a different doctrinal route, arguing that the attribution of the adverse cyber operation to the State of origin turns that State into a "sanctuary State." *See* Sklerov, *supra* note 7, at 72.

168. Corfu Channel, *supra* note 78, at 22.

cyber operations resulting in transboundary harm. Consequently, the victim-State has demonstrated that the State of origin has knowingly allowed its territory to be used contrary to the rights of other States.

The resumption of adverse cyber operations shifts the burden of proof to the State of origin, raising the issue of whether the State of origin has acted with due diligence. To demonstrate compliance with its due diligence obligation, it must establish convincingly that it took all reasonable measures to prevent further adverse cyber operations committed by use of the private IT system in question.[169] Whether the State of origin's preventive measures were reasonable depends on the circumstances of the individual case.[170] The answer depends not just on the technical and economic feasibility of preventive measures and the State's capacities,[171] but also on human rights considerations.[172] Measures aimed at preventing adverse cyber operations may conflict with international human rights, including the freedom of speech and of the press, freedom of access to information, freedom of correspondence and telecommunication, freedom to conduct a business, or protection of personal data.[173] Thus, there may be limitations or restrictions on a State's ability to undertake preventive measures when to do so would infringe upon these rights.

If adverse cyber operations are launched from private IT systems that have not been identified as original sources of such operations, it is, in general, impossible to argue that the State of origin knowingly allowed its territory to be used for the commission of acts harmful to another State. Moreover, when a State's human rights obligations are considered, it cannot be presumed that the State of origin "constantly kept a close watch over"[174] all

169. *See* Ney & Zimmermann, *supra* note 41, at 64; TALLINN MANUAL 2.0, *supra* note 7, at 46–47.

170. For an in-depth analysis, see TALLINN MANUAL 2.0, *supra* note 7, at 47–50; *see also* Ney & Zimmermann, *supra* note 41, at 63–64.

171. This point leads to the problem that the standard of due diligence may fluctuate depending on States with different capacities. *See, e.g.*, Kolb, *supra* note 145, at 123; August Reinisch & Markus Beham, *Mitigating Risks: Inter-State Due Diligence Obligations in Case of Harmful Cyber Incidents and Malicious Cyber Activity – Obligations of the Transit State*, 58 GERMAN YEARBOOK OF INTERNATIONAL LAW 101, 101–06 (2015).

172. *See* Dörr, *supra* note 10, at 95–96; Ney & Zimmermann, *supra* note 41, at 64; Reinisch & Beham, *supra* note 171, at 112; Schmitt, *In Defense of Due Diligence*, *supra* note 23, at 74–75.

173. Concerning the human rights dimension of cyberspace governance, see Ney & Zimmermann, *supra* note 41, at 58; *see also* TALLINN MANUAL 2.0, *supra* note 7, at 179–208.

174. Corfu Channel, *supra* note 78, at 18.

private IT systems within its territory, or had the possibility of so doing.[175] Such a presumption assumes the State knew, or should have known, of the abuse, or potential abuse, of all private systems within its territory: an assumption that is unrealistic in the cyber world.[176] In this regard, the *Tallinn Manual* 2.0 has correctly pointed to the "difficulty of mounting comprehensive and effective defences against all possible cyber threats."[177]

IV. CONCLUSION

The foregoing analysis leads to the following conclusions:

- Adverse cyber operations are those operations that employ cyber capabilities with the objective of causing harm in, or by the use of, cyberspace.
- In responding to adverse cyber operations, victim-States may have recourse to their right to act in self-defense or to take countermeasures. Both acts in self-defense and countermeasures may be directed, however, only against such States of origin as are internationally responsible for the particular cyber operation. Accordingly, the operation must be attributable to those States.

---

175. *See* Dörr, *supra* note 10, at 95; Walter, *Obligations of States*, *supra* note 128, at 76; *see also* TALLINN MANUAL 2.0, *supra* note 7, at 45 (noting that the International Group of Experts rejected the view that a State is "required to monitor cyber activities on its territory").

176. *See also* Walter, *Obligations of States*, *supra* note 128, at 74 (stating that the "Corfu Channel formula['s] . . . requirement of 'knowledge' of the harmful [cyber] activity will only be met in rare cases"). For a contrary view, see Richard Garnett & Paul Clarke, *Cyberterrorism: A New Challenge for International Law*, *in* ENFORCING INTERNATIONAL LAW NORMS AGAINST TERRORISM 465, 479 (Andrea Bianchi ed., 2014) (arguing that "in a situation where there have been repeated instances of hostile computer activity emanating from a State's territory directed against another State, it seems reasonable to presume that the host State had knowledge of such attacks and so should incur responsibility"); *see also* Sklerov, *supra* note 7, at 13. Sklerov asserts, "repeated failure by a state to take criminal action against its attackers will result in it being declared a sanctuary state, allowing victim-states to use active defenses against cyberattacks originating from within its borders." *Id.* at 72.

177. TALLINN MANUAL 2.0, *supra* note 7, at 45 (noting that the International Group of Experts concluded that "it would be unreasonable to assert that an obligation of prevention exists in the cyber context" since "[s]uch a requirement would impose an undue burden on States"). For similar perspective, see Dörr, *supra* note 10, at 95; Walter, *Obligations of States*, *supra* note 128, at 78.

- The victim-State bears the burden of proof for those facts that, if true, establish that the operation was caused (1) by conduct (2) of a person, group of persons, or entity (3) attributable to the State of origin.
- The question of whether evidence gathered and submitted by the victim-State is sufficient to discharge its burden of proof is decided in light of the applicable standard of proof. The standard of proof applies as a function of the gravity of the alleged breach of international law. Accordingly, the standard of proof applicable in the individual case is to be derived from a sliding scale. The standard of proof thus identified also applies with regard to attribution and causality.
- With regard to acts of self-defense, in general, a high standard of proof, i.e., convincing evidence applies.
- For the taking of countermeasures, either a high or an intermediate standard of proof, i.e., either convincing evidence or preponderance of the evidence may apply.
- Even when convincing evidence is the standard, the burden of proof may be discharged solely based on indirect evidence.
- If the victim-State is able to present evidence tracing the adverse cyber operation to a State IT system, it can be reasonably inferred that a State organ initiated the operation. Once established, the State of origin may then prove the negative fact that a State organ had not authored the operation. It discharges its burden of proof if it is able to prove that it took all reasonable measures to prevent the launching of adverse cyber operations from its IT systems.
- If the victim-State is able to establish only that the adverse cyber operation can be traced to a private IT system within the territory of the State of origin, it may be unable to prove the factual prerequisites for attributing the operation to the State of origin. In order to prevent an inability to sustain the burden of proof and the standard of proof from becoming a *probatio diabolica*, the State of origin is under a duty to cooperate in good faith with the victim-State. This duty is an expression of the general due diligence obligation found within *Corfu Channel* for States "not to knowingly allow its territory to be used for acts contrary to the rights of other States." If the State of origin is unwilling to cooperate or cooperates only *mala fide*, the victim-State is allowed to infer from such behavior that the operation of the private IT system at the time the adverse cyber operation was launched was attributable to the State of origin.

- Once the State of origin has been notified that a private IT system located within its territory has been the source of adverse cyber operations, for any subsequent operations emanating from the same IT system it may be inferred that the State of origin knew, or should have known, the private IT system was used for cyber operations possibly resulting in transboundary harm. It follows that the burden of proof shifts to the State of origin, which may discharge its burden of proof if it proves that it acted in accordance with its due diligence obligation, i.e., the *Corfu Channel* rule.
- If adverse cyber operations are activated through private IT systems that had never been identified as original sources of adverse cyber operations, it is, in general, impossible to draw the inference that the State of origin failed to comply with its due diligence obligations because it cannot be presumed that the State of origin constantly monitored, or could have monitored, all private IT systems within its territory.

The aforementioned conclusions form a nuanced and balanced approach concerning evidentiary issues surrounding causality and attribution for adverse cyber operations. The resulting concept accounts for the legitimate interests of both the victim-State and the State of origin without besting either of the two. A victim-State need not fear that its hands are tied, whereas a State of origin need not fear that it will be unduly and prematurely exposed to measures of self-defense or countermeasures.