"Hacking Back" by States and the Uneasy Place of Necessity within the Rule of Law

Henning Lahmann[*]

I. Introduction 453
II. Hacking Back as Strategy 455
III. Hacking Back and the Rule of Law in Cyberspace 457
1. Vulnerabilities Exploitation and the Rule of Law 458
2. Justifying Hack Backs and the Problem of Attribution 461
IV. Invoking Necessity 464
1. Status and Preconditions of Necessity to Justify Hack Backs 464
2. State of Necessity and the Rule of Law 467
V. Upholding the Rule of Law: Context-Specific Rules for Recurring Emergency Situations 470
1. A *lex specialis* Emergency Regime for Cyberspace 472
2. Possible Elements of an Emergency Regime for Cyberspace 473
VI. Concluding Remarks 476

I. Introduction

Over the past few years, malicious cyber security incidents have become an ever more pressing issue due to a higher frequency of attacks targeting increasingly sensitive and high-stakes assets. In particular, critical infrastructures,[1] such as the telecommunications or energy sector, seem ever more often to be the focus of advanced persistent threats[2] or other malicious

[*] Senior Researcher, Digital Society Institute, ESMT Berlin; <henning.lahmann@esmt.org>; I want to thank the participants of the 2019 ESIL Research Forum "The Rule of Law in Cyberspace" in Göttingen for their helpful comments on the first draft, especially Dr. *Irene Couzigou*, Dr. *Paulina Starski*, Dr. *Andreas Kulick*, and Prof. *Nicholas Tsagourias*.

[1] Though there is no uniform, internationally recognized official definition of the term "critical infrastructures", the definition given by the U.S. Department of Homeland Security provides some helpful guidance: "[…] infrastructure whose assets, systems, and networks, whether physical or virtual, are considered so vital to the [state] that their incapacitation or destruction would have a debilitating effect on security, national economic security, national public health or safety, or any combination thereof"; see <https://www.dhs.gov> (all URLs accessed on 30.6.2019).

[2] "An advanced persistent threat (APT) uses continuous, clandestine, and sophisticated hacking techniques to gain access to a system and remain inside for a prolonged period of time, with potentially destructive consequences"; Kaspersky Lab, What Is an Advanced Persistent Threat (APT)?, <https://www.kaspersky.com>.

cyber operations.[3] Large industrial facilities that serve as the central supplier for entire regions have become the targets of confirmed cyber-attacks in the past years.[4] A successful malicious cyber operation on such infrastructures has potentially devastating effects, including the safety of the affected state's civilian population, as could be witnessed in 2015 when an attack crippled significant parts of the Ukrainian power grid.[5] Virtually all experts agree that the problem of critical infrastructure protection against malicious cyber operations is only going to become more urgent in the near future.[6]

Faced with such threats from cyberspace, policymakers across the globe have started viewing purely passive, defensive measures as too often insufficient.[7] Instead, official cybersecurity strategies have gradually shifted towards what is commonly known as "active cyber defense", understood as measures to stop or mitigate malicious cyber operations outside of the defender's systems. As proposals aiming at implementing such capabilities proliferate among a growing number of states, it is thus high time to assess their potential ramifications for the global cybersecurity environment and the enforcement of international law in cyberspace. Specifically analyzing "hack backs" as the most frequently invoked variation of active cyber defenses, the present paper argues that such policies threaten to undermine the already fragile rule of law in cyberspace, and they do so in two distinct ways.

After explicating the notion of "hacking back" and the implementation of respective policies by states, the concept of the rule of law is briefly sketched out. Subsequently, it is shown how the technical requirement to rely on vulnerabilities in the target system's soft- or hardware in order to perform hack backs means that state security agencies have a strong incentive to refrain from disclosing found vulnerabilities. It is shown how this practice, by design, weakens the rule of law in cyberspace.

---

[3] See German Federal Office of Information Security, Die Lage der IT-Sicherheit in Deutschland, 2018, 10.

[4] See the example provided in the Cisco 2018 Annual Cybersecurity Report, 36, <https://www.cisco.com>.

[5] *R. M. Lee/M. J. Assante/T. Conway*, Analysis of the Cyber Attack on the Ukrainian Power Grid, Electricity Information Sharing and Analysis Center, 2016, <https://ics.sans.org>; another, potentially even more dangerous security incident followed in December 2016, see *A. Greenberg*, "Crash Override": The Malware that Took Down a Power Grid, Wired, 2017, <https://www.wired.com>.

[6] *A. Carcano*, Critical Infrastructure Under Persistent Threat, Security Magazine, 2018, <https://www.securitymagazine.com>.

[7] *G. Mascolo/R. Steinke*, BND könnte Lizenz zum "Hack Back" bekommen, 2018, <https://www.sueddeutsche.de>.

The second, more crucial way in which hacking back policies undermine the rule of law is found on a more fundamental level. First, the paper explains how hack backs would be justifiable under international law in principle if it were not for the pervasive problem of timely attribution in cyberspace. Consequently, the subsequent section explores how recourse to a state of necessity seems to lend itself as a feasible way out of the attribution dilemma. The paper argues that while many experts consider necessity applicable to situations of a cyber emergency, the doctrine as found in customary international law presents a problem for the rule of law.

Tackling both challenges, the final section outlines possibilities to operationalize a conventional emergency regime for cyberspace that does not ignore the vulnerability disclosure problem.

II. Hacking Back as Strategy

As malicious cyber operations have become more prolific, more states have begun considering cybersecurity strategies that comprise hacking back capabilities. The 2018 Cyber Strategy of the United States (U.S.) Department of Defense, for instance, explicitly endorses the concept of "defending forward",[8] even though the contours of the doctrine remain somewhat vague.[9] In Germany, a number of politicians have expressed the desire to create the legal basis for hacking back operations,[10] which has resulted in an ongoing political debate.[11] Furthest ahead, however, are the Swiss. In 2017, the Confederation enacted a law that, for the first time, explicitly provides for the Federal Intelligence Service a right to hack back in certain circumstances.[12]

---

[8] U.S. Department of Defense, Cyber Strategy 2018, 1, <https://media.defense.gov>: "We will defend forward to disrupt or halt malicious cyber activity at its source, including activity that falls below the level of armed conflict."

[9] *D. Weinstein*, The Pentagon's New Cyber Strategy: Defend Forward, Lawfare, 2018, <https://www.lawfareblog.com>.

[10] Westfälische Nachrichten, Kanzleramtschef: Prüfen Möglichkeit von Cyber-Gegenangriffen, 18.3.2018, <https://www.wn.de>.

[11] See, e.g., BT-Drs. 19/2645, 11.6.2018; BT-Drs. 19/5472, 5.11.2018.

[12] Art. 37 para. 1 of the Swiss Federal Law on the Intelligence Service: "If computer systems and computer networks that are located abroad are employed for the purpose of attacking critical infrastructures in Switzerland, the Federal Intelligence Service (Nachrichtendienst des Bundes) is permitted to infiltrate these computer systems and computer networks in order *to disrupt, thwart, or slow down the access to information*. The Federal Council decides on the execution of such a measure." (Translation by the author, emphasis added).

For the purpose of this examination, it is important to note that the understanding of the concept is not clear-cut. For example, the Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations defines "active cyber defence" vaguely as "the taking of proactive measures outside the defended cyber infrastructure".[13] "Hack backs" as a subset of active cyber defense are described as actions "against an identified source of a malicious cyber operation designed to mitigate the effects of, or stop, the malicious activity, or to gather technical evidence that can be used for attribution purposes".[14]

To give another example from recent state practice, the German Federal Ministry of the Interior attempts to conceive hacking back capabilities as part of a more comprehensive cyber defense strategy.[15] To this end, it has developed a five-stage model of escalating responses to threats emanating from cyberspace. While the first two stages envisage purely defensive measures such as firewalls, anti-virus software, or diverting the attacker's data in order to avert damage, the third – gathering evidence for the purpose of identifying the perpetrator – approaches the threshold of what the Tallinn Manual considers "hacking back".[16]

Stages 4 and 5 of the official German strategy refer to "hack backs" more narrowly and in line with how the concept is usually discussed in international legal scholarship. In order to mitigate a successful attack's consequences, the second-highest stage allows for accessing the attacking systems with the aim of deleting stolen data, a step that had first been considered after foreign agents had hacked the secure network of the German Federal Parliament in 2015.[17] As *ultima ratio*, finally, defenders shall get the permission to physically destroy servers that, for instance, serve as command and control systems for an ongoing malicious operation. That would, for example, be electronically possible by implanting malware that triggers the malfunctioning of the attacker system's cooling elements, causing components to melt down due to overheating.[18]

---

[13] *M. Schmitt* (ed.), Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, 2017, Glossary.

[14] *M. Schmitt* (note 13).

[15] *G. Mascolo/R. Steinke* (note 7).

[16] *T. Reinhold/M. Schulze*, Digitale Gegenangriffe, Arbeitspapier, Stiftung Wissenschaft und Politik, 2017, 7, <https://www.swp-berlin.org>.

[17] As "stolen" data can easily be copied, the effectiveness of such measures is very much in doubt, see *J. Diehl/F. Reinbold*, Wenn der Staat zum Hacker wird, Spiegel Online (2017), <http://www.spiegel.de>.

[18] *M. Schmitt* (note 13), 20.

For a legal assessment, it is important to note that not all hacking back operations need a justification under international law, especially not if one assumes that there is no stand-alone rule of respect for sovereignty below the "coercion" threshold of the prohibition of intervention.[19] For the purpose of the following examination, the focus will be on hack backs that (1) are carried out by or on behalf of a state; (2) aim at fending off an ongoing malicious cyber operation against interests of the state or at alleviating the consequences of a malicious cyber operating; and (3) infringe on legally protected interests of another state. For the sake of the argument, it will be assumed that even without amounting to "coercion" within the meaning of the prohibition of intervention, hacking back operations are capable of violating the territorial sovereignty of the target state and thus require a circumstance precluding wrongfulness, as otherwise the acting state would be internationally responsible for the violation of sovereignty. Finally, it should be noted that the subsequent arguments principally concern non-automated hack backs, i.e., conduct that is carried out and controlled by human agents. While the outlined legal principles generally also apply to automated hacking back operations, which are on the agenda of a number of international actors, the aspect of automation adds further legal questions that are not the subject of this article.

III. Hacking Back and the Rule of Law in Cyberspace

Having defined hacking back and its implementation as an integral element of an increasing number of states' official cybersecurity strategies, it is now to be examined how such policies potentially have an impact on the notoriously fickle rule of law in cyberspace.[20] Of course, it must first be determined how the rule of law is to be understood in this context, as the concept itself is somewhat elusive:

[19] See, e.g., the official position of the UK to this effect: "Sovereignty is of course fundamental to the international rules-based system. But I am not persuaded that we can currently extrapolate from that general principle a specific rule or additional prohibition for cyber activity beyond that of a prohibited intervention.", *J. Wright*, Cyber and International Law in the 21st Century, 2018, <https://www.gov.uk>; this standpoint has met with criticism in the literature, see only *J. Biller/M. Schmitt*, Un-Caging the Bear? A Case Study in Cyber Opinio Juris and Unintended Consequences, EJIL Talk, 2018, <https://www.ejiltalk.org>.

[20] A useful overview of the various issues is provided by *K. Giles*, Prospects for the Rule of Law in Cyberspace, Strategic Studies Institute, 2017, <https://apps.dtic.mil>.

> "while all (or almost all) agree that the Rule of Law is an indispensable aspect of a worthwhile legal system, there is less agreement on the content and scope of the ideal".[21]

This general problem concerning the exact outlines of the concept is exacerbated when it comes to its application to the system of international law, which lacks centralized enforcement. Therefore, a more modest, minimal understanding of the "rule of law" seems pertinent: While the concept can be conceived as comprising a great number of characteristics, its most crucial aspects for the question at hand are the stability, consistency, and predictability of the norms that govern a community[22] – here, the community of states and all other subjects of international law. Without the establishment of these principles, the norms are unable to fulfill one of their principal functions, which is to provide legal certainty, or, in the words of sociologist *Niklas Luhmann*, the counterfactual stabilization of expectations.[23] In the absence of legal certainty, a legal system is arbitrary.[24] If such a state of affairs persists, the addressees of a legal system will eventually become unable to even form expectations.[25]

As already hinted at, it is submitted that due to both technical and doctrinal considerations, cybersecurity policies that comprise hacking back as part of their toolbox pose a challenge to the rule of law in cyberspace, especially concerning the elements of stability and predictability, potentially contributing to its further erosion.

1. Vulnerabilities Exploitation and the Rule of Law

To understand the first challenge, which concerns the technical aspects of the policy, it is important to note that by definition, hacking "back" is still

---

21 *G. Lamond*, The Rule of Law, in: A. Marmor (ed.), The Routledge Companion to Philosophy of Law, 2012, 495.

22 Stanford Encyclopedia of Philosophy, The Rule of Law, 2016, <https://plato.stanford.edu>; while this article limits the understanding to these three relevant aspects, note that *L. Fuller* identifies eight requirements for the rule of law in his influential work Morality of Law, 1969, 39; see on this in detail *C. Murphy*, Lon Fuller and the Moral Value of the Rule of Law, Law and Philosophy 24 (2005), 239 et seq.; *J. Raz*, The Authority of Law: Essays on Law and Morality, 1979, 210 et seq., provides a similar list of features that comprises variants of above three elements of the rule of law.

23 *N. Luhmann*, Law as a Social System, 2004, 147.

24 *A. Zwitter*, The Rule of Law in Times of Crisis: A Legal Theory on the State of Emergency in the Liberal Democracy, University of Groningen Faculty of Law Research Paper Series No. 10 (2013), 23.

25 Stanford Encyclopedia (note 22).

hacking. This means that in order to retain the capability to engage in defensive cyber operations in the case of a cyber-attack, states must participate in the practice of holding back found vulnerabilities in soft- and hardware.[26] Vulnerabilities can be defined as

> "weaknesses of software, hardware, or online services that can be used to damage the confidentiality, integrity, or availability of those systems or the data they store".[27]

They can be divided into zero-day exploits and n-day exploits; vulnerabilities of the former category are not yet known to the public and, more crucially, the actor capable of providing a patch.[28] The existence of such errors in virtually every piece of published software is an inherent and inescapable fact of coding. By not disclosing those weaknesses, the states' responsible security agencies are then able to utilize them to gain access to the systems of attackers or suspects. In order to do that, security agencies use the vulnerabilities as components of malware that is able to intrude a target's systems and execute the desired operations, such as extracting data or obstructing the operations of the hacked machine. Similar methods are employed for surveillance purposes in criminal investigations or intelligence operations.[29] There are exceptions, to be sure – not all active defensive measures in cyberspace require the exploitation of software flaws. However, retaining the capability to hack back as a strategy will necessitate the build-up of an "arsenal" of offensive software tools that can be employed against malicious actors if necessary. Evidently, however, the practice inevitably results in an overall less secure cyberspace environment. For the stability of the legal system governing the internet, and thus for the rule of law, this has serious ramifications.

For one, the respective policies to that end directly undermine emerging legislation that aims at increasing the security level of information and communications technology (ICT) devices and applications. For example, the new European Union (EU) Cybersecurity Act envisages as one of its core components a scheme to certify that

---

[26] *M. Baumgärtner/M. Gebauer/M. Knobbe/M. Rosenbach*, Wer will die Verantwortung übernehmen, Unschuldige zu töten?, Der Spiegel 35 (2018).

[27] National Telecommunications and Information Administration, Vulnerability Disclosure Attitudes and Actions, 2016, 3, <https://www.ntia.doc.gov>.

[28] *S. Herpig*, Governmental Vulnerability Assessment and Management, Stiftung Neue Verantwortung, 2018, 7, <https://www.stiftung-nv.de>.

[29] *R. Pfefferkorn*, Security Risks of Government Hacking, The Center for Internet and Society, 2018, 5, <https://cyberlaw.stanford.edu>.

> "an ICT product, process or service has no known vulnerabilities at the time of the certification's release and that it complies with international standards and technical specifications".[30]

According to the European lawmakers in charge of the proposal, such a certification will prove, *inter alia*, "that processes are in place to identify all known vulnerabilities and deal with any new ones", and "that products, processes or services are designed to be secure and that they are fitted with up-to-date software *without known vulnerabilities*".[31] The legislation thus aims at providing trust in the digital environment in order to underpin, by technical means, the overall legal stability of cyberspace. If, at the same time, intelligence services or law enforcement agencies are allowed to abstain from disclosing vulnerabilities that they gain knowledge of, the authority of the EU certification process is gradually being chipped away: Just imagine what would happen to the credibility of the certification scheme if the undisclosed vulnerability of a previously certified product or service leads to widespread financial loss within the European economy due to a severe malicious cyber security incident made possible by an exploitation of that vulnerability. A state-sanctioned policy to retain vulnerabilities thus, in a way, contradicts current legislative efforts on the level of the European Union. The ensuing loss of trust in the security of the network environment affects the stability of the system as a whole. While this does not in itself affect the stability of the rule of law – as the norms remain, on paper, unharmed – it has a negative effect on the norms' capability to stabilize expectations.

What is more, there is a certain degree of arbitrariness immanent in most current state behavior concerning vulnerabilities. If a state engages in a practice that factually decreases ICT security, the least it should be obligated to do as an accompanying measure is to implement some process aiming at providing accountability. Without a formalized policy in place that enables democratic control and transparency to a certain extent,[32] from the perspective of citizens using ICT products and services, the practice will necessarily remain opaque. Without such a process, hacking back policies are neither predictable nor impartial, two core features of the rule of law as circum-

30 European Parliament, Cybersecurity Act: Build Trust in Digital Technologies, Press Release, 10.7.2018, <http://www.europarl.europa.eu>.

31 European Parliament (note 30), (emphasis added).

32 Addressing this issue, the United States implemented a vulnerabilities equities process in 2017, see Vulnerabilities Equities Policy and Process for the United States Government, 15.11.2017, 1, <https://www.whitehouse.gov>; however, the policy has been subject to ongoing criticism, see *L. Hay Newman*, Feds Explain Their Software Bug Stash – But Don't Erase Concerns, Wired, 2017, <https://www.wired.com>.

scribed above. Perhaps even more consequential is the fact that the practice advances insecurity more generally by enabling malicious actors to break the law: The longer a certain vulnerability remains undisclosed, the likelier it gets that cyber criminals or officials from adversarial states discover the same weakness, allowing them to exploit it for unlawful activity.[33]

2. Justifying Hack Backs and the Problem of Attribution

A further, more fundamental challenge posed by such cybersecurity strategies relates to the issue of justifying hacking backs under international law in the case that the conduct outside of the acting state's territorial boundaries infringes upon the legally protected interests of another state. While, as mentioned, that does not necessarily need to be the case,[34] it is to be assumed that the "hack back" will frequently constitute at least an infringement upon the target state's sovereignty – provided it is accepted as a stand-alone rule of customary international law. Those authors who advocate for such a right hold that a state's sovereignty is affected, *inter alia*, by a cyber operation that disrupts the information and telecommunication infrastructure in the target state to a more than neglectable degree.[35] If a "hack back" interferes with normal operations of the ICT infrastructure or even physically destroys infrastructure on the territory of the target state, the latter's sovereignty would thus be infringed. Under certain circumstances, the hack back may amount to an intervention or even a use of force. Each of these cases triggers the need to invoke a circumstance precluding the wrongfulness of the act in accordance with customary international law as reflected in the 2001 International Law Commission (ILC) Articles on State Responsibility (ASR).

The legal qualification of the hack back thereby principally depends on the consequences of the conduct. As soon as the "scale and effects" of the measures are equivalent to those of an operation carried out with kinetic means rising to the level of a use of force pursuant to Art. 2(4) of the United Nations (UN) Charter, they can only be justified as self-defense in accord-

[33] See *B. Sander*, Recommendation 13(e), in: United Nations Office for Disarmament Affairs (ed.), Voluntary, Non-Binding Norms for Responsible State Behaviour in the Use of Information and Communications Technology, 2017, 145.

[34] See the discussion in *M. Schmitt* (note 13), 17 et seq.; a hack back for mere reconnaissance purposes, without the intention of gathering evidence for judicial ends, would arguably not infringe upon the sovereignty of the target state.

[35] See only *M. Schmitt/L. Vihul*, Respect for Sovereignty in Cyberspace, Tex. L. Rev. 95 (2017), 1639.

ance with Art. 51 of the UN Charter, which also serves as a circumstance precluding wrongfulness more generally in accordance with Art. 21 ASR.[36] That is the case if the hacking back operation leads to the death or bodily harm of persons or the significant destruction of objects in the target state. The recourse to self-defense, of course, is only available when the hack back is carried out as a necessary and proportionate measure against an armed attack of the other state.

If the defensive cyber operation's consequences stay below the use of force threshold yet qualify either as an intervention or at least as an infringement upon the target state's sovereignty, the hack back needs to rely on the right to conduct countermeasures pursuant to Arts. 49 to 54 ASR in order to be justified.[37] Without going too much into detail at this point, hack backs as countermeasures undertaken to halt or thwart a malicious cyber operation need to be proportional and may not violate fundamental human rights, amount to a use of force, or otherwise be in contradiction to a peremptory norm of international law.

More crucially, however, the cyber security incident that the hacking back operation is directed against must be attributed to the state that is affected by the hack back. In the case of self-defense, the targeted state needs to be actually responsible for the malicious operation itself that the defensive measure is a reaction to.[38] When the defending state intends to justify its hacking back to the source as a countermeasure, it will arguably be sufficient if the target state had been in breach of its due diligence obligation to prevent malicious cyber operations emanating from its own territory.[39]

Much has been written about the attribution problem in cyberspace which does not need to be repeated here.[40] Suffice it to add that this article does not claim that it will never be possible for states to successfully identi-

---

[36] On the generally accepted "scale and effects" criterion see *M. Schmitt* (note 13), Rule 69.

[37] *M. Roscini*, World Wide Warfare – Jus Ad Bellum and the Use of Cyber Force, Max Planck UNYB 14 (2010), 113.

[38] Unless, according to the "unwilling or unable" doctrine, the state is either directly responsible for a use of force by private actors emanating from its territory if it was unwilling or unable to prevent it, or it at least is under the obligation to tolerate defensive measures against the private actors on its territory; the customary status of this legal construct remains unclear, and an in-depth discussion is beyond the scope of this article; see only *A. Deeks*, "Unwilling or Unable": Toward a Normative Framework for Extraterritorial Self-Defense, Va. J. Int'l L. 52 (2012), 483.

[39] *O. Gross*, Cyber Responsibility to Protect: Legal Obligations of States Directly Affected by Cyber-Incidents, Cornell Int'l L. J. 48 (2015), 481, 494.

[40] See only *M. Roscini*, Evidentiary Issues in International Disputes Related to State Responsibility for Cyber Operations, in: J. D. Ohlin/C. Finkelstein/K. Govern (eds.), Cyberwar: Law & Ethics for Virtual Conflicts, 2015, 215.

fy the source of a malicious operation and hence establish responsibility.[41] However, for a hack back to be both effective and justified in the above manner, it is crucial that attribution is successful in a timely manner, as self-defense and countermeasures alike require the conduct in question to be *necessary*. That will usually not be the case if the original attack constitutes an already properly concluded, past event. Yet, reliable attribution of cyber security incidents to a malicious actor requires a thorough, multi-source investigation that does not merely rely on forensic, easily falsified digital evidence. In this regard, the basic rule is that the more time and resources are at the defending state's disposal, the more evidence will be gathered so the more likely reliable identification and attribution of the incident's authors become: "The quality of attribution is a function of the available time", in the words of *Rid* and *Buchanan*.[42] To identify the source of a cyber security incident and to establish the necessary connection to a state quickly enough will, for this reason, remain a considerable, structural obstacle for the time being.[43] In time-sensitive situations, as for instance during on-going malicious cyber operations against critical infrastructures, the probability to be able to identify the responsible actor decreases dramatically.[44]

If timely attribution proves impossible in a given case, self-defense and countermeasures not only fail as justifications for hacking back operations. To be sure, such lack of accountability poses a more general problem for the rule of law in cyberspace. The attribution problem, which is a direct result of the technical peculiarities of cyberspace, makes plausible deniability of state conduct in cyberspace possible. That being the case, malicious actors are enabled to evade the need to justify their conduct by means of invoking legal language. As a consequence, the conduct ultimately falls outside the bounds of law.[45] Or, put more bluntly in the words of *Jack Goldsmith*, "anonymity is a norm destroyer".[46]

---

41 See on this issue generally *B. Schneier*, Click Here to Kill Everybody, 2018, 54 et seq.

42 *M. Schulze*, Hacking Back? Technische und politische Implikationen digitaler Gegenschläge, SWP-Aktuell, 2017, 2, <https://www.swp-berlin.org>; *T. Rid/B. Buchanan*, Attributing Cyber Attacks, Journal of Strategic Studies 38 (2015), 4, 32.

43 *M. Dickow*, Stellungnahme zur Öffentlichen Anhörung des Verteidigungsausschusses des Deutschen Bundestages am 22. Februar 2016, Ausschussdrucksache 18(12)640.

44 *M. Dickow* (note 43).

45 See *S. Ratner*, Persuading to Comply: On the Deployment and Avoidance of Legal Argumentation, in: J. L. Dunoff/M. A. Pollack (eds.), Interdisciplinary Perspectives on International Law and International Relations, 2013, 568, 585.

46 *J. Goldsmith*, Cybersecurity Treaties. A Skeptical View, 2011, 12, <http://media.hoover.org>.

IV. Invoking Necessity

Then, if a state finds itself in a perilous situation caused by a malicious cyber operation that calls for an immediate response by way of conducting a hack back against the source that can be identified but not sufficiently linked to a state actor, one obvious path to justify its conduct is the invocation of a state of necessity. The Tallinn Manual, for one, pragmatically acknowledges the legal possibility of states to

> "act pursuant to the plea of necessity in response to acts that present a grave and imminent peril […] to an essential interest when doing so is the sole means of safeguarding it".[47]

Thus, the Manual effectively reiterates the wording of Art. 25 ASR. And even though it points to the high legal threshold for the invocation of necessity,[48] it explicitly suggests the remedy as a possible justification for conducting hack backs in the case that attribution of a significant malicious cyber operation targeting the state's critical infrastructure proves impossible.[49] However, while necessity thus lends itself as a seemingly compelling alternative in light of the technical intricacies of cyberspace, the doctrine comes with its own, specific set of issues, in particular its relation to the rule of law. This will be laid out and assessed in the following.

1. Status and Preconditions of Necessity to Justify Hack Backs

Even though some authors claim that the status of the necessity doctrine under customary international law remains unsettled and is still somewhat controversial,[50] for quite some time, there has been a growing body of international jurisprudence that acknowledges the existence of such a reme-

[47] *M. Schmitt* (note 13), 135.
[48] *M. Schmitt* (note 13), 135.
[49] *M. Schmitt* (note 13), 138.
[50] See *R. D. Sloane*, On the Use and Abuse of Necessity in the Law of State Responsibility, AJIL 106 (2012), 447, 450 et seq.; *J. Kurtz*, Adjudging the Exceptional at International Investment Law: Security, Public Order and Financial Crisis, ICLQ 59 (2010), 325, 344.

dy.[51] The leading case in this regard is without doubt the 1997 International Court of Justice decision on the *Gabčíkovo-Nagymaros Project*.[52]

Whatever its exact status, the nature of general necessity as an absolute exception meant for rare emergency situations is reflected both by the negative phrasing of Art. 25 ASR[53] and by the strict preconditions that aim at setting up a deliberately high threshold for ever successfully taking recourse to it. The rule states that states may rely on necessity to preclude the wrongfulness of a certain act only if there is no other way to protect an essential interest against a grave and imminent peril and if the act does not seriously impair an essential interest of another state or the international community as a whole. Moreover, necessity is unavailable if either the international obligation that is sought to be suspended excludes its invocation or if the state itself has contributed to the situation of necessity.

Applying the norm to "hack backs" taken in order to stop or mitigate the consequences of malicious cyber operations, experts are generally in agreement that at least a state's critical infrastructures, as understood in the U.S. definition above, can be considered essential interests as required by the provision.[54] A cyber security incident that threatens the functioning of those infrastructures is furthermore potentially a grave and imminent peril within the meaning of Art. 25 ASR. While both requirements are inherently context-dependent, a peril can be conceived as grave once a level of severity is reached that makes the significant impairment or even destruction of the essential interest likely.[55] As for imminence, the authoritative commentary to the ASR asserts that the peril needs to be "objectively established and not merely apprehended as possible".[56] While this addresses the timespan *before* a malicious cyber operation and thus relates to the question from what point a hack "back" could be considered lawful under necessity, for example when the acting state has gained knowledge of an impending operation, the more relevant issue for the context at hand is arguably the permissibility of hacking back operations *after* the initial discovery of a cyber security incident. Once a malicious cyber operation is completed, the peril will usually not be "imminent" anymore. However, most hack backs will only be tech-

[51] On necessity generally *A. Tanzi*, Necessity, State of, in: R. Wolfrum (ed.), MPEPIL, 2012; *J. D. Ohlin/L. May*, Necessity in International Law, 2016.

[52] *Gabčíkovo-Nagymaros Project* (*Hungary v. Slovakia*) (Judgment), ICJ Reports 1997, para. 51.

[53] International Law Commission (ILC), Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries, 2001, Art. 25, para. 14.

[54] *M. Schmitt* (note 13), 135.

[55] *M. Agius*, The Invocation of Necessity in International Law, NILR 56 (2009), 95, 103.

[56] ILC Articles, Commentaries (note 53), Art. 25, para. 15.

nically useful if the operation is in some sense still ongoing anyway, for example when the intrusive malware that imperils the functioning of the affected critical infrastructure continues to be operated remotely via a command-and-control server. As long as this part of the cyber operation has not ceased, the peril should qualify as imminent. At the same time, a peril should additionally be accepted as imminent if the initial cyber operation has triggered consequences that are continuing to threaten essential interests. This would be the case, for example, if highly classified, national security-sensitive information was extracted and the affected state launches a hack back measure with the aim of retrieving (i.e., copying and/or deleting) the data from the adversary's (or a third party's) servers.

Still trickier is the additional precondition that the hack back must be the only way to address the situation. History suggests that this is where most invocations of necessity ultimately fail.[57] As the "only means" requirement aspires to ensure that necessity will only be available as a truly last resort, a state that is faced with a serious cyber security incident is under the obligation to exhaust all purely defensive, non-intrusive ways to avert the peril before it may invoke necessity in order to preclude the wrongfulness of infringing upon the rights of another state as a consequence of a hacking back operation.[58] Whether this is realistic within the cyber security context is principally a technical question. However, many official strategies, for instance the above mentioned five-stage model of the German Federal Ministry of the Interior, point in the right direction in that they envisage hacking back as the *ultima ratio*, to be undertaken only once purely defensive measures have proven insufficient in the case at hand.

Finally, it bears repeating that given the architecture of cyberspace, the consequences of a hack back always risk crossing the use of force threshold pursuant to Art. 2(4) UN Charter – for instance, if the attackers had used the cyber infrastructure of a hospital in order to cover their tracks and the defender's operation then destroys a hospital server that not only controls the attackers' moves but also functions that are critical for the patients'

[57] See, e.g., *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory* (Advisory Opinion), ICJ Reports 2004, para. 140; *M/V "SAIGA" (No. 2)* (*Saint Vincent and the Grenadines v. Guinea*), ITLOS Reports 1999, para. 135; *CMS Gas Transmission Company v. Argentine Republic*, ICSID Case No. ARB/01/8, 2005, paras. 323 et seq.; *Sempra Energy International v. Argentine Republic*, ICSID Case No. ARB/02/16, 2007, para. 350.

[58] Likewise *M. Schmitt* (note 13), 139.

well-being.[59] Contrary to some persisting voices in academic literature,[60] a state's use of force can never be justified by way of invoking necessity. The reason for this is doctrinal and rooted in the systematic layout of contemporary international law: The rules on the use of force as enshrined in the UN Charter constitute a special regime in relation to the more general necessity rule as far as forceful conduct by states is concerned.[61] The actual use of force thus triggers the application of the special regime and, at the same time, precludes a state from relying on necessity – in accordance with Arts. 25(2)(a) and 55 ASR – for the purpose of justifying said conduct.[62] To be sure, this does not necessarily exclude all acts that can be defined as "force", as forceful conduct below the threshold of Art. 2(4) UN Charter is arguably at least conceivable.[63] Such "force" could for example be used in the course of police or other law enforcement measures, possibly even if taken outside of the acting state's jurisdiction.[64]

2. State of Necessity and the Rule of Law

The foregoing section has shown that a successful recourse to the customary state of necessity for the purpose of justifying a hacking back operation that infringes upon the rights of another state may rarely be successful given the remedy's deliberately high legal threshold, yet it can nevertheless not be ruled out in principle. And indeed, as long as timely and reliable at-

[59] This example is, even if overused, still instructive to illustrate the potential ramifications of hacking back, see *M. Baumgärtner/M. Gebauer/M. Knobbe/M. Rosenbach/W. Wiedmann-Schmidt*, Hacker mit Dienstausweis, Der Spiegel 48 (2017), 31.

[60] See, e.g., *Schmitt* (note 13), 140.

[61] On special regimes, sometimes misleadingly called self-contained regimes, as particularly strong manifestations of the *lex specialis* principle see in particular *M. Koskenniemi*, Fragmentation of International Law: Difficulties Arising from the Diversification and Expansion of International Law. Report of the Study Group of the International Law Commission, United Nations 2006, A/CN.4/L.682, paras. 124, 152; *B. Simma/D. Pulkowski*, Of Planets and the Universe: Self-Contained Regimes in International Law, EJIL 17 (2006), 483.

[62] See most succinctly *O. Corten*, L'état de Nécessité Peut-Il Justifier Un Recours À La Force Non Constitutif D'agression?, The Global Community Yearbook of International Law & Jurisprudence 4 (2004), 11, 48: "Une interprétation du texte de la Charte, tel qu'il a été conçu puis interprété par le biais de plusieurs résolutions adoptées par l'Assemblée générale, confirme que la prohibition du recours à la force représente un régime juridique qui *n'admet pas d'échappatoire*." (emphasis added).

[63] See, e.g., *O. Corten*, The Law Against War, 2010, 85; Institut de Droit International, Session de Santiago – 2007, 10A Resolution EN, 27.10.2007, para. 5.

[64] *O. Corten* (note 62), 860; see on this, e.g., *Case Concerning the Detention of Three Ukrainian Naval Vessels* (*Ukraine v. Russian Federation*) (Request for the Prescription of Provisional Measures) (Order), ITLOS Case No. 26 (2019), paras. 73 et seq.

tribution of cyber security incidents continues to be a problem, and the more prevalent malicious cyber operations against critical infrastructures become, it does not at all seem implausible that the temptation to invoke necessity will only grow in the near future. This, it is argued, is bound to become a considerable problem for the rule of law in cyberspace.

The state of necessity shares its legal-theoretical origins with, and is structurally closely related to, the state of exception.[65] While the latter concerns the state's legal relationship to its citizens, the former pertains to the relations between subjects on the same legal plane – natural persons or, on the level of international law, states. Contrary to the well-known assertions by theorists such as *Carl Schmitt*[66] or *Giorgio Agamben*,[67] the situation triggered by such emergency regimes does not as such fall outside the bounds of law.[68] This does not mean, however – in view of their ambiguous and indistinct character – that necessity and the state of exception do not confront legal theory with rather intricate questions. As "irregular grounds"[69] for precluding the wrongfulness of an act, they serve as a reaction to an emergency situation that is claimed to be "beyond the boundaries of the normal operation of the legal regime concerned".[70] In order to address the emergency, they thus suspend the *normal* operation of the law.[71]

---

65 *G. Agamben*, State of Exception, 2005, 24: "A recurrent opinion posits the concept of necessity as the foundation of the state of exception"; also see *W. Vázquez Irizarry*, Exception and Necessity: The Possibility of a General Theory of Emergency, <https://law.yale.edu>, transl. of: Excepción y necesidad: la posibilidad de una teoría general de la emergencia, Sela 2010, 274.

66 *C. Schmitt*, Political Theology. Four Chapters on the Concept of Sovereignty, 2005, 13: "There exists no norm that is applicable to chaos. For a legal order to make sense, a *normal situation* must exist […]" (emphasis added).

67 *G. Agamben* (note 65), 39: "The state of exception is an anomic space in which what is at stake is a force of law".

68 *T. Stahlberg/H. Lahmann*, A Paradigm of Prevention: Humpty Dumpty, the War on Terror, and the Power of Preventive Detention in the United States, Israel, and Europe, The Am. J. Comp. L. 59 (2011), 1051, 1085; *D. Dyzenhaus*, Schmitt v. Dicey: Are States of Emergency Inside or Outside the Legal Order?, Cardozo L. Rev. 27 (2006), 2005.

69 *W. Küper*, Von Kant zu Hegel. Das Legitimationsproblem des rechtfertigenden Notstandes und die freiheitsphilosophischen Notrechtslehren, JZ 60 (2005), 105; this observation led to a reluctance in some jurisdictions to codify necessity into (criminal) law. In Germany, for example, justifying necessity was incorporated into the criminal code only in 1975, finally losing its extra-statutory existence, see *W. Küper*, Grundsatzfragen der "Differenzierung" zwischen Rechtfertigung und Entschuldigung, JuS 27 (1987), 81.

70 *T. Gazzini/W. G. Werner/I. F. Dekker*, Necessity Across International Law: An Introduction, NYIL 41 (2010), 3, 8 et seq.

71 According to *Hans Kelsen*, the state of emergency thus acts as a *lex specialis* to normal law, see *A. Zwitter* (note 24), 15.

By definition, such an *exceptional* recourse must be limited to truly unforeseen circumstances and furthermore inherently temporarily limited. Invoking necessity legitimates a state to act outside of its normatively expected performance. Such exceptional legal mechanisms imply that they must be terminated once the cause for their implementation has come to an end, and that the return to normality should be accomplished as soon as possible.[72] Routine, frequent reliance on emergency provisions or a prolonged state of emergency, on the other hand, will eventually inverse the relationship between rule and exception and thus undercut the rule of law, making the system of law in operation less stable and predictable.[73] Contra *Agamben*, the state of exception may not be an anomic space, but its exceptionality entails that it operates on a different set of premises that ought not translate to the level of "normal" law. Overuse of an emergency or necessity provision, then, must eventually lead to the creeping normalization of the exception, thereby slowly eroding – and possibly even superseding[74] – the supposed operation of the legal system that the provision itself is a part of.

Considering the developments in transnational cybersecurity in recent years and acknowledging the unique features of cyberspace that cause the attribution problem, it is unlikely that cyber emergencies that prompt states to engage in hacking back measures will remain isolated or rare occurrences. The emerging threats that states face are well-documented and recur on an almost regular basis. The inability to attribute a malicious cyber operation quickly enough and with the sufficient degree of legal certainty, as has been shown, is a function of the structural layout of cyberspace. It is, therefore, anything but "exceptional". Yet, invoking necessity implies the existence of unforeseen, truly exceptional circumstances. A general, indistinct necessity norm that operates in such a way as to suspend a state's obligations under international law is thus inherently conceptually insufficient to deal with the problem at hand. That does not imply that Art. 25 ASR could never be

72 *A. Zwitter* (note 24), 9 et seq.

73 See, e.g., the assessment of France's state of emergency since 2015, *B. Boutin/C. Paulussen*, From the Bataclan to Nice: A Critique of France's State of Emergency Regime, ASSER Policy Brief No. 2016-01, 2016, <http://ssrn.com/abstract=2811602>; a comparison with the law of occupation, which is also an extra-ordinary body of law that is inherently time-limited, and the ramifications of prolonged occupations is equally instructive, see *O. Ben-Naftali/A. Gross/K. Michaeli*, Illegal Occupation: Framing the Occupied Palestinian Territory, Berkeley J. Int'l L. 23 (2005), 551, 606: "Structurally, the law of occupation bears strong resemblance to an emergency regime. […] [A] situation of emergency is separated and distinguished from the ordinary state of affairs as it signifies an occurrence which does not conform to the rule. Because the emergency situation is the exception, its duration must be limited and it must generate no permanent effects; it merely suspends the rule."

74 Stanford Encyclopedia (note 22).

successfully invoked by a state that is confronted with a serious cyber security incident. Yet, the fact that the likeliness for cyber emergencies to become a recurring problem is reasonably high means that the necessity defense as rooted in customary international law should not be used as the default legal basis for hacking back policies. Instead, the international community should consider adapting the general concept of necessity to the specific context in order to enable states to conduct emergency measures that might affect other states' legally protected interests while at the same time clearly circumscribing the requirements and limits of such conduct. As will be shown in the next section, there is historic precedent for such a step.

V. Upholding the Rule of Law: Context-Specific Rules for Recurring Emergency Situations

On 18.3.1967, the government of the United Kingdom (UK) found itself in a desperate situation. Just off the Cornish coast, the Liberian supertanker *Torrey Canyon*, carrying over 100,000 tons of crude oil, had ran aground.[75] The accident caused a massive oil spill that endangered the English Channel coast. After some futile attempts to control the impending disaster, the UK sent fighter jets with the mission to bomb vessel and leaking oil in order to burn it before it could reach the beaches.[76] Giving the command, the responsible authorities were aware that this conduct would lead to the destruction of a foreign ship on the high seas.[77] Despite legal controversy and palpable discomfort among the community of states at the time, the case has become a frequently cited textbook example of a successful invocation of the customary necessity defense.[78]

In the aftermath of the incident, however, the international community quickly arrived at the conclusion that in the dawning age of supertankers, it was rather unlikely that the casualty would remain an isolated event.[79]

---

[75] See for a recount of the event BBC, On This Day: 1967: Supertanker Torrey Canyon Hits Rocks, BBC News (1967), <http://news.bbc.co.uk>; *J. E. Smith* (ed.), "Torrey Canyon" Pollution and Marine Life, 1968.

[76] See BBC, On This Day: 1967: Bombs Rain down on Torrey Canyon, BBC News (1967), <http://news.bbc.co.uk>.

[77] *A. E. Utton*, Protective Measures and the "Torrey Canyon", Boston College Industrial and Commercial Law Review 9 (1968), 613, 625.

[78] See only ILC Articles, Commentaries (note 53), Art. 25, para. 9.

[79] See *D. M. Dzidzornu/B. M. Tsamenyi*, Enhancing International Control of Vessel-Source Oil Pollution Under the Law of the Sea Convention, 1982: A Reassessment, U. Tas. L. Rev. 10 (1991), 269, 278 et seq.

Quite the contrary, "such incidents might recur at any time".[80] Thus, invoking the exact rationale against basing the reaction to a potentially frequent event on the uncertain and perilous concept of the general state of necessity, the international community recognized the need for more detailed and comprehensive rules capable of governing a distressed state's immediate response to an oil hazard caused by a shipwreck[81] – rules, in other words, that could *tame* and indeed *normalize* the state of necessity.[82] After some contentious deliberations, the result was the Convention Relating to Intervention on the High Seas in Case of Oil Pollution Casualties (Intervention Convention), drafted by the legal committee of the Intergovernmental Maritime Consultative Organization (IMCO) and adopted on 29.11.1969.[83] For the first time, states had written rules that precisely provided for the preconditions of coastal state intervention and the measures that were deemed permissible.[84]

It bears emphasizing, however, that not all conceivable emergency scenarios are amenable to being regulated by clear and precise specific rules. Granting wide-reaching emergency powers in exceptionally perilous situations remains a delicate balancing act, and some situations defy obvious legalistic solutions. The inherently ambiguous place of the state of necessity within the (constitutional) rule of law raised to the surface during the German Federal Constitutional Court's deliberations on Section 14(3) of the 2005 Aviation Security Act, which provided the state with the authority to shoot down passenger planes in the case that they had been hijacked by terrorists in a 9/11-type scenario. Even in such a dire situation that could lead to thousands of deaths, the Court considered it impossible to legally conceptualize a right to suspend the state's duty to not by itself threaten the life

---

[80] *R. Ago*, Eighth Report on State Responsibility. Addendum: The Internationally Wrongful Act of the State, Source of International Responsibility (Part I) (concluded), United Nations 1980, A/CN.4/318/ADD.5-7, para. 36.

[81] *R. M. M'Gonigle/M. W. Zacher*, Pollution Politics and International Law, 1979, 145.

[82] *W. Vázquez Irizarry* (note 65), 2; one of *Carl Schmitt's* main arguments against the legalization of the state of exception was his assertion that "[g]eneral norms […] cannot anticipate the myriad factual scenarios a state might confront within public emergencies or the measures necessary to deal with them", see *E. J. Criddle/E. Fox-Decent*, Human Rights, Emergencies, and the Rule of Law, HRQ 34 (2012), 39, 42 et seq.; while one may concede that it is factually impossible to foresee *all* possible emergency scenarios, it is argued that to the degree that some potentially recur frequently, the enactment of specific rules in relation to such cases is both possible and expedient.

[83] *C. C. Emanuelli*, The Right of Intervention of Coastal States on the High Seas in Cases of Pollution Casualties, U.N.B.L.J. 25 (1976), 79.

[84] Since its coming into force in 1994, the right to intervention has found a second home in Art. 221(1) of the United Nations Convention on the Law of the Sea. The Intervention Convention remains applicable.

of the passengers on the plane, be they doomed either way or not.[85] Thus, by implication, it held that such a scenario cannot adequately be represented within the legal order. An executive's decision to shoot down the plane could therefore merely rely on an extra-statutory state of necessity, which might exculpate the individual actors involved,[86] but which could in no case render the official act itself lawful, i.e., exculpate the state. In the words of *Ernst Burgbacher*, Member of the German Federal Parliament during the legislative debate, "there are clashes of rights that evade exact legislative description".[87] While this example concerns the situation of a state's internal constitutional law, the underlying legal principles and the implications for the rule of law more generally are applicable to the context of international law as well.

Considering malicious cyber operations, however, it is submitted that the situation is more closely related to events the Intervention Convention sought to regulate than the shooting down of a hijacked passenger plane. Therefore, the following section outlines a special emergency regime for cyber security incidents.

1. A *lex specialis* Emergency Regime for Cyberspace

The principal distinction between emergency scenarios that underlie the Intervention Convention and those that the German legislators imagined to regulate with Section 14(3) of the Aviation Security Act is the fact that the latter situation directly affects innocent individuals and their human dignity,[88] a principle recognized in international human rights law.[89] Even if the safety of persons is imperiled in the case of a cyber security incident that is answered with a hack back – potentially by either operation – the stakes are obviously lower than in the scenario concerning the shooting down of a passenger jet. Therefore, it is suggested, a specific emergency regime for cyberspace, modelled on the example of the Intervention Convention, that clearly prescribes the preconditions and legal consequences of engaging in hacking back or other active cyber defense measures in order to stop or mitigate cyber security incidents and that acts as a *lex specialis* framework in

85 BVerfG, 1 BvR 357/05, 15.2.2006.

86 The Court explicitly abstained from the question of criminal culpability, see BVerfG (note 85), 130.

87 Bundestag, Plenary Protocol 15/89, 20.1.2004, 7888.

88 BVerfG (note 85), 121.

89 See the Preamble and Art. 1 of the Universal Declaration of Human Rights and the Preamble of the International Covenant on Civil and Political Rights.

relation to the customary state of necessity could alleviate the inherent tensions between necessity and the rule of law.[90]

Of course, any such regime would still need to put a special, explicit emphasis on the protection of human rights. As has been observed by several scholars in view of states in distress,

> "[t]he same political pressures that prompt states to declare states of emergency also generate strong incentives for states to violate their human rights obligations during emergencies".[91]

The same holds true for invoking a state of necessity under international law, which is, in contrast to the declaration of an emergency which addresses the state's own constitutional order and thus the relationship to its citizens, outward-looking, i.e., addressing the state's relationship to other states. But in this regard, too, a special legal framework should take the well-being of uninvolved individuals into account. After all, the precarious status of human rights protections during emergencies lies at the heart of the "immense pressure"[92] that such situations exert on the rule of law.

2. Possible Elements of an Emergency Regime for Cyberspace

Having established the rationale for a specific emergency regime for cyberspace, the penultimate section of the examination shall briefly sketch out some possible elements of such a legal framework, drawing on some of the principles derived from the Intervention Convention.[93]

First of all, an operative cyber emergency regime needs to comprise precise and workable definitions of key concepts such as "cyber security incident" or "critical infrastructures", serving as specifications to the "grave and imminent peril" and "essential interest" requirements of the customary necessity norm. The respective notions in the Intervention Convention are "pollution or threat of pollution of the sea by oil, following upon a mari-

---

[90] The relation between such a special emergency regime and Art. 25 ASR is governed by Art. 55 ASR: "These articles do not apply where and to the extent that the conditions for the existence of an internationally wrongful act or the content or implementation of the international responsibility of a State are governed by special rules of international law." See on this in detail *M. Koskenniemi* (note 61), paras. 46-222.

[91] *E. J. Criddle/E. Fox-Decent* (note 82), 45 et seq.

[92] *A. Zwitter* (note 24), 4.

[93] See in more detail *H. Lahmann*, Unilateral Remedies to Cyber Operations. Self-Defence, Countermeasures, Necessity, and the Question of Attribution, 2020, 272 et seq.

time casualty or acts related to such a casualty" and "coastline or related interests". Furthermore, the Intervention Convention makes a distinction between situations of "normal emergency" and "extreme urgency". If an incident does not pose an imminent threat, certain procedural safeguards are required before the distressed state is allowed to act, such as the duty to inform all states and other stakeholders whose rights might be affected by the envisaged hack back. Such a rule seems appropriate for cyber emergencies as well. If a state faced with a cyber security incident, for example, intends to engage in hacking back measures against a privately-owned server on foreign territory, both the operator of the system and the territorial state should be informed prior to conducting the operation in order to give them the opportunity to terminate the perilous activity themselves, if possible. Only in a situation of extreme urgency, or when the notified actors prove incapable or unwilling to undertake the necessary measures, would the distressed state then be permitted to go ahead and hack into the server.[94]

The Intervention Convention also acknowledges that an emergency situation on the high sea inherently affects a multi-stakeholder environment. Actions taken by the imperiled state must therefore consider the interests not just of other states but also of private actors such as ship-owners, trading companies, or insurance firms, by being proportionate towards all of them. This arrangement perfectly reflects the situation in cyberspace with its countless private actors such as service and infrastructure providers or telecommunications companies. Their legal interests should be taken into account by the state that intends to engage in measures against a cyber security incident. At the same time, given the interconnectedness of cyberspace and as a result the vast number of potentially affected third parties, not every stakeholder can reasonably be taken into account. Therefore, only such interests should be considered legally relevant that will foreseeably and directly be negatively affected by the hacking back operation, for example a private operator of the targeted server.

Perhaps most strikingly, the Intervention Convention separates the distressed state's need to act without delay from questions of accountability

[94] In this regard, the 2007 cyber-attack against Estonia may illustrate the point. After the Estonian authorities had determined that the bulk of malicious activity during the persistent DDoS attacks originated from servers located on Russian territory, they requested the governmental agencies in Moscow to initiate measures to stop the attacks against Estonia. Russia, however, ignored the call for help. Under the emergency model proposed here, the Estonian authorities would then have had the right to (electronically) intervene on Russian territory in order to attempt to interrupt the attacking data flow; see *G. Keizer*, Estonia Blamed Russia for Backing 2007 Cyberattacks, Says Leaked Cable, Computerworld (2010), <https://www.computerworld.com>.

and instead calls for neutral and formalized *ex post facto* assessment of the incident by means of conciliation or arbitration. In particular, the question of fault or wrongdoing on the part of the distressed state is hence adjourned. This model could also be applied to cyber security incidents. As has been pointed out, an imperiled state will likely not have enough time to reliably establish attribution prior to conducting a hack back as an emergency measure, yet, that should not preclude it from acting against the source of a security incident in the case that no other, less intrusive measures are feasible. If, after the fact, an independent arbiter determines that the state erroneously targeted an entirely uninvolved system or acted disproportionately, it will be liable to compensation. At least as far as the question of responsibility for the initial malicious cyber operation is concerned, and not the question of proportionality or the determination of inflicted damage, this might even be an appropriate role for the kind of neutral and international "attribution councils" that have repeatedly been proposed by a number of scholars and non-governmental organizations.[95] Instead of attempting to allocate responsibility *ex ante*, which is likely doomed to fail, such institutions could instead serve as impartial arbiters in an incident's aftermath. However, it should be added that so far, states have been rather skeptical towards the idea of establishing an independent body tasked with attributing cybersecurity incidents, so it remains to be seen whether they might reach an agreement to that effect in the future. Nevertheless, *ex post facto* assessment can of course be achieved in other ways, for instance by the international courts and tribunals, arbitration, or *ad hoc* fact-finding commissions.

Finally, an emergency regime for cyberspace that provides for the possibility to engage in hacking back measures should take into account the potential ramifications of states' offensive cyber capabilities by including a rule that addresses the issue of soft- and hardware vulnerabilities, as touched upon above. To be sure, that does not call for an in any way internationalized vulnerabilities equities process. Such an obligation would be impossible to implement due to national security considerations. However, the legal framework could provide that states that build up hacking back capabilities by retaining vulnerabilities have a duty to put some legal process in place that follows principles of the rule of law, ensuring that the policy is applied

[95] See, e.g., the "Global Cyber Attribution Consortium" proposed by the RAND Corporation, <https://www.rand.org>, or Microsoft's initiative to establish an IAEA-type body, see *S. Waterman*, Microsoft Calls for UN-Type Body to Attribute Big Cyberattacks, fedscoop (2016), <https://www.fedscoop.com>.

consistently.[96] Such a process could take the shape of an official vulnerabilities management policy, as has been installed, for example, by the United States and the United Kingdom,[97] but other modalities such as parliamentary or joint executive control are conceivable as well. States should be free to find a suitable arrangement. In any case, the default standard should be a duty to disclosure: This means that found soft- or hardware flaws should be communicated to the vendor without undue delay in order to enable the company to patch its product and by that contribute to the overall security of cyberspace, "unless there is a specific, justifiable reason for retaining and using them in law enforcement, intelligence or military programs".[98]

VI. Concluding Remarks

The article has attempted to expose and assess some of the challenges to the rule of law in cyberspace posed by states' emerging policies regarding "active cyber defenses" and particularly "hacking back" measures as their most important sub-category. In view of the pervasive attribution dilemma, intrusive reactions to cyber security incidents without reliable attribution will become more likely. In such a situation, recourse to the state of necessity could become more and more frequent, which would eventually undermine the rule of law. As a solution to the problem, a specific emergency regime for cyber security incidents has been suggested.

To date, it has proven difficult to agree on new rules that could regulate state conduct in cyberspace.[99] Therefore, it is of course hard to tell how likely it may be that states could come up with such a legal framework. Then again, written rules concerning the law of the sea seemed nearly impossible for a long time as well.

Last, one potential reservation against the proposed emergency regime is that its enactment might mean that hacking back operations become more

[96] See on this *N. Tsagourias*, Recommendation 13(j), in: United Nations Office for Disarmament Affairs (ed.), Voluntary, Non-Binding Norms for Responsible State Behaviour in the Use of Information and Communications Technology, 2017, 241 et seq.

[97] Vulnerabilities Equities Policy and Process (note 32); see also the recently announced UK Equities Process, conducted by the Government Communications Headquarter and the National Cyber Security Centre, <https://www.gchq.gov.uk>: "The starting position is always that disclosing a vulnerability is in the national interest." From a rule of law standpoint, the main issue with such processes is the (deliberate and explained) lack of transparency concerning the decision-making.

[98] *S. Herpig* (note 28), 18.

[99] See only *J. Goldsmith* (note 46); *A. Segal/M. C. Waxman*, Why a Cybersecurity Treaty Is a Pipe Dream, CNN.com (2011), <http://globalpublicsquare.blogs.cnn.com>.

frequent and likely. That is a reasonable critique which demands further scrutiny, not least as hack backs have a number of potentially grave downsides – such conduct might rather easily hit the wrong target, endangering or even harming uninvolved individuals or negatively affecting another state's infrastructures, for example. Moreover, hitting back through cyberspace inherently bears a significant risk of escalation that is not easily curbed and might lead to a serious inter-state conflict. Although the Intervention Convention, for one, has so far never been invoked,[100] laws that intend to regulate the exception tend to facilitate recourse to unnecessary or harmful emergency measures, as could recently be witnessed in the United States.[101] Rules that provide for the exception, including a right to hack back, should only be invoked as rarely as in any way possible. After all, the fundamental tension between the rule of law and the state of exception that lies at the heart of this matter has so far not been resolved.

[100] *P. Wendel*, State Responsibility for Interference with the Freedom of Navigation in Public International Law, 2007, 49.

[101] See *P. H. Schuck*, The Real Problem With Trump's National Emergency Plan, The New York Times (2019), <https://www.nytimes.com>.