Limiting viral spread: automated cyber operations and the principles of distinction and discrimination in the Grey Zone

Kaminska, M.K.; Broeders, D.W.J.; Cristiano, F.; Jančárková, T.; Lindström, L.; Visky, G.; Zotz, P.

2021 13th International Conference on Cyber Conflict
Going Viral
T. Jančárková, L. Lindström, G. Visky, P. Zotz (Eds.)
2021 © NATO CCDCOE Publications, Tallinn

Limiting Viral Spread: Automated Cyber Operations and the Principles of Distinction and Discrimination in the Grey Zone

Monica Kaminska
Postdoctoral Researcher
The Hague Program for Cyber Norms
Institute of Security and Global Affairs
Leiden University
The Hague, The Netherlands
m.k.kaminska@fgga.leidenuniv.nl

Fabio Cristiano
Postdoctoral Researcher
The Hague Program for Cyber Norms
Institute of Security and Global Affairs
Leiden University
The Hague, The Netherlands
f.cristiano@fgga.leidenuniv.nl

Dennis Broeders
Full Professor of Global Security and Technology
The Hague Program for Cyber Norms
Institute of Security and Global Affairs
Leiden University
The Hague, The Netherlands
d.w.j.broeders@fgga.leidenuniv.nl

Abstract: The fact that States resort to automated cyber operations like NotPetya, which spread virally and have indiscriminate effects, raises the question of how the use of these might be regulated. As automated operations have thus far fallen below the threshold of the use of force, the letter of international humanitarian law (IHL) does not provide such regulation. In IHL, the principles of distinction and discrimination hold that attacks should in their targeting distinguish between the civilian population and combatants, and between civilian objects and military objectives. Attacks must not be indiscriminate, and operations that might foreseeably spread to affect civilian objects are prohibited. This paper draws inspiration from the legal principles of distinction and discrimination to suggest a non-binding norm for responsible State behaviour with regard to automated operations that fall below the threshold of the use of force: the norm proposes that States should design cyber operations so as to prevent them from indiscriminately inflicting damage. The paper finds that in the case of automated

cyber operations, a distinction between the nature of the operation and the use of the operation does not make sense because the design (nature) of the malware defines the use. In order to conform with the norm, responsible States should conduct a review of cyber operations prior to their execution. Finally, as the paper illustrates with a comparative analysis of NotPetya and Stuxnet, the post-incident forensic analysis of an operation can allow third parties and victims to determine whether the operation's designer conformed with the norm. This can help set a normative benchmark by providing a basis upon which States may call out unacceptable behaviour.

Keywords: *automated cyber attacks, international humanitarian law, indiscriminate attacks, cyber norms, sub-threshold operations*

1. INTRODUCTION

Automated State-led cyber operations have the potential to spread and affect systems uncontrollably. The WannaCry and NotPetya attacks of 2017 are the most pressing examples of operations that were not designed to limit harmful effects on systems, which meant that they were able to destroy data on networks supporting a wide range of services, from national healthcare to international commercial shipping. Meanwhile, existing legal frameworks, particularly international humanitarian law (IHL), are insufficient to regulate conduct with reference to attacks like WannaCry and NotPetya that fall below the threshold of the use of force.[1] In this paper, drawing inspiration from IHL, we propose a new norm against indiscriminate cyber operations below the threshold of the use of force. The norm holds that States should design cyber operations so as to prevent them from indiscriminately inflicting damage. While the norm draws inspiration from IHL, it deviates from IHL in that it does not distinguish between lawful and unlawful objects as categories. Instead, any operation that does not seek to target a malware's payload at a particular system; that is, lacks any form of distinction and discrimination, would be considered a violation of the norm.

In considering how we might borrow from the ideas of legal weapons review and targeting law in the context of regulating automated cyber operations, we find that such operations challenge the classic IHL distinction between the 'nature' and 'use' of weapons. In order to conform to the norm, we argue that responsible State actors should conduct a normative review of cyber operations at the design stage to ensure

1 While the initial NotPetya attack was launched in the context of an armed conflict between Russia and Ukraine, the malware spread globally and inflicted most of its damage outside Ukraine. The operation itself fell below the threshold of the use of force as it did not cause physical injury or significant damage beyond economic and data losses. For a longer discussion on the application of international law to NotPetya see: Michael Schmitt and Jeffrey Biller, 'The NotPetya Cyber Operation as a Case Study of International Law', EJIL: Talk! (blog), 11 July 2017, https://www.ejiltalk.org/the-notpetya-cyber-operation-as-a-case-study-of-international-law/.

that the operations are designed to limit harmful effects. This recommendation stems directly from the existing recognition in the scholarly literature that cyber weapons are not 'inherently indiscriminate' and can be designed so as to accomplish the perpetrator's goals without causing significant damage beyond the intended target.[2]

The paper is divided into three sections. First, referring particularly to recent attribution statements and State contributions to the Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security (OEWG), we argue that States are starting to get worried about automated cyber attacks, which indicates the need for the development of a norm against indiscriminate sub-threshold operations. Second, we discuss why a distinction between 'nature' and 'use' does not make sense in the context of automated cyber operations and propose that responsible States should conduct a 'normative' review of the design of cyber operations prior to their launch. Third, we compare and contrast two well-known cyber operations, NotPetya and Stuxnet, to show how a post-incident analysis of an operation can reveal whether the attacker sought to limit the operation's uncontrolled harmful effects.

2. THE NEED FOR A NORM TO LIMIT AUTOMATED ATTACKS BELOW THE THRESHOLD OF AN ARMED ATTACK

NotPetya, and WannaCry before it, forced States to think about the nature and the permissibility of automated cyber attacks below the threshold of armed conflict. The financial and operational damage done, and the indiscriminate way in which the malware spread, set these attacks apart. When a number of States coordinated their attributions of the NotPetya attack to Russia, the US and the UK made references to its automated nature. The UK condemned 'its indiscriminate design' that caused it to spread beyond its primary Ukrainian targets.[3] The United States called it out in light of the ongoing conflict between Russian and Ukraine but also underlined that 'this was … a reckless and indiscriminate cyber attack that will be met with international consequences'.[4] The unofficial American condemnation was a lot harsher. Tom Bossert, President Trump's homeland security advisor, was adamant that a spoken or unspoken red line around how the United States expects fellow countries to behave on the internet had been violated: 'The United States thinks any malware that propagates recklessly, without bounds, violates every standard and expectation of proportionality

2 Steven M. Bellovin, Susan Landau, and Herbert S. Lin, 'Limiting the Undesired Impact of Cyber Weapons: Technical Requirements and Policy Implications', *Journal of Cybersecurity* 3, no. 1 (2017): 61.

3 Foreign and Commonwealth Office, 'Foreign Office Minister Condemns Russia for NotPetya Attacks', GOV.UK, 15 February 2018, https://www.gov.uk/government/news/foreign-office-minister-condemns-russia-for-notpetya-attacks.

4 White House, 'Statement from the Press Secretary', 15 February 2018, https://trumpwhitehouse.archives.gov/briefings-statements/statement-press-secretary-25/.

and discrimination. Truly responsible nations do not behave this way.'[5] However, given that attacks like NotPetya take place below the threshold of the use of force – or are at least not called out by States as a use of force – the principles of IHL do not apply. In other words, there is no easy resort to principles of discrimination and proportionality to judge an indiscriminate and viral attack below the threshold.

State worries about indiscriminate cyber attacks have also surfaced in the recent and ongoing rounds of the UN processes on determining responsible State behaviour in cyberspace. The OEWG wrapped up its deliberations with a report in March 2021 and the parallel process of the UN Group of Governmental Experts (UN GGE) is still yet to be finalised.[6] States can submit their contributions in writing to the OEWG and, in contrast to the closed diplomatic process of the UN GGE, have them published on the OEWG website.[7] In January 2020, Switzerland voiced its concerns: 'While the majority of cyber operations have so far been executed in a precise and targeted manner from a technical point of view, we have recently seen cases within which cyber tools were used at random and causing unintended harmful effects.'[8] Both the first[9] and the second[10] pre-drafts of the report included an unchanged reference to this problem in the threat section: 'Pursuit of increasing automation and autonomy in ICT operations was also put forward as a specific concern.' In their responses to the first draft report, States like Brazil, Ecuador and the Netherlands explicitly supported the inclusion of this concern, the latter adding that '[t]hese independently operating and developing cyber operations are, once launched, outside the control of the initiators, and therefore the adherence to the framework of responsible behaviour including

5 Cited in: Andy Greenberg, *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers* (New York: Doubleday, 2019), 244.

6 For some background on these processes see: Tim Maurer, 'A Dose of Realism: The Contestation and Politics of Cyber Norms', *Hague Journal of the Rule of Law* (2019): 1–23; Dennis Broeders and Bibi van den Berg, 'Governing Cyberspace. Behavior, Power, and Diplomacy', in *Governing Cyberspace. Behavior, Power, and Diplomacy*, eds. Dennis Broeders and Bibi van den Berg (London: Rowman and Littlefield, 2020), 1–15; Dennis Broeders (2021) 'The (im)possibilities of addressing election interference and the public core of the internet in the UN GGE and OEWG: a mid-process assessment', *Journal of Cyber Policy*, forthcoming.

7 UNODA, 'Open-Ended Working Group', accessed 25 December 2020, https://www.un.org/disarmament/open-ended-working-group/.

8 Federal Department of Foreign Affairs FDFA et al., 'Position Paper on Switzerland's Participation in the 2019-2020 UN Open-Ended Working Group on "Developments in the Field of Information and Telecommunications in the Context of International Security" and the 2019-2021 UN Group of Governmental Experts on "Advancing Responsible State Behavior in Cyberspace in the Context of International Security"', January 2020, https://unoda-web.s3.amazonaws.com/wp-content/uploads/2020/02/switzerland-position-paper-oewg-gge-final.pdf.

9 'Initial "Pre-Draft" of the Report of the OEWG on Developments in the Field of Information and Telecommunications in the Context of International Security', n.d., https://unoda-web.s3.amazonaws.com/wp-content/uploads/2020/03/200311-Pre-Draft-OEWG-ICT.pdf.

10 'Second "Pre-Draft" of the Report of the OEWG on Developments in the Field of Information and Telecommunications in the Context of International Security', n.d., https://front.un-arm.org/wp-content/uploads/2020/05/200527-oewg-ict-revised-pre-draft.pdf.

international law cannot be ensured'.[11] As the final OEWG report[12] only reflects consensus opinions, the reference to indiscriminate cyber attacks was dropped there and moved to the Chair's summary.[13] This document contains issues that were put forward by multiple states but did not achieve consensus and will be discussed further in coming iterations of the UN processes on responsible behaviour in cyberspace.

At this point two things need to be disentangled. First, there is a conflation of automation and autonomy. While these are partly overlapping concepts, we focus in this paper on automation rather than autonomy. Autonomy is most fiercely debated in the context of Lethal Autonomous Weapon Systems (LAWS), where a whole range of ethical and legal questions are raised on the issue of (the lack of) human control and computer autonomy in military weapons and systems.[14] The debate on artificial intelligence (AI) enabled cyber attacks also touches on the issue of autonomy, as AI could enable malware to react autonomously to changing circumstances and possibilities. This debate is relatively overhyped: for most attackers, AI is not needed as the available cyber automation techniques serve their purposes.[15] This paper focuses on the automated, viral quality of cyber attacks like NotPetya and the way they spread indiscriminately. Second, if States flag automation as a problem, one of the next questions is whether this can be addressed by international law or by non-binding norms (or both). As indicated above, no State has formally stated that NotPetya violated any principles of international law, let alone IHL. Even though NotPetya seemed 'most poised to burst out of the grey zone between war and peace', State reactions indicate that it did not.[16] However, there have been some efforts to develop non-binding norms to acknowledge and address the problem of automated cyber attacks. The 2018 *Paris Call for Trust and Security in Cyberspace* explicitly acknowledges the emergence of 'malicious cyber activities in peacetime' that are 'threatening or resulting in significant, indiscriminate or systemic harm to individuals

11 *The Kingdom of the Netherlands' Response to the Pre-Draft Report of the OEWG*, n.d., https://front.un-arm.org/wp-content/uploads/2020/04/kingdom-of-the-netherlands-response-pre-draft-oewg.pdf.

12 United Nations General Assembly, 'Open-Ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security Final Substantive Report' (United Nations, 10 March 2021), https://front.un-arm.org/wp-content/uploads/2021/03/Final-report-A-AC.290-2021-CRP.2.pdf.

13 United Nations General Assembly, 'Open-ended Working Group on Developments in the Field of Information and Telecommunications in the Context of International Security Third substantive session 8–12 March 2021 Chair's Summary' (United Nations, 10 March 2021), https://front.un-arm.org/wp-content/uploads/2021/03/Chairs-Summary-A-AC.290-2021-CRP.3-technical-reissue.pdf.

14 Michael C. Horowitz, 'The Ethics & Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons', *American Academy of Arts & Sciences* 145, no. 4 (Fall 2016): 25–36; Kenneth Anderson and Matthew C. Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation under International Law', in *The Oxford Handbook of Law, Regulation and Technology* (Oxford: Oxford University Press, 2017), https://doi.org/10.1093/oxfordhb/9780199680832.001.0001; Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W.W. Norton & Company, n.d.).

15 Ben Buchanan et al., 'Automating Cyber Attacks' (Washington, D.C.: Center for Security and Emerging Technology, November 2020).

16 Ben Buchanan, *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics* (Cambridge, MA: Harvard University Press, 2020), 302.

and critical infrastructure' and 'welcome[s] calls for their improved protection'.[17] The Global Commission on the Stability of Cyberspace covers large-scale automated attacks by asserting that 'state and non-state actors should not commandeer the general public's ICT resources for use as botnets or for similar purposes'.[18] This norm seems primarily focused on the use of botnets, but the 'similar purposes' clause might be applicable to automated attacks like NotPetya.

In this paper we argue that viral, automated attacks could be addressed by a non-binding norm for responsible State behaviour below the threshold of the use of force that draws inspiration from legal principles derived from IHL. Norms have been constructed in this way before. Some of the eleven norms in the 2015 UN GGE report are reiterations of legal principles – such as the principle of due diligence or respect for human rights law – indicating that norms and laws are perhaps more of a continuum than a strict dichotomy.[19] Inspired by the principles of distinction and discrimination in IHL, this norm would bar indiscriminate cyber operations below the threshold of the use of force. First, under IHL, the legality of a weapon (system) is among other things determined by the fact that the weapon system cannot be indiscriminate by nature. This rule refers to the 'nature of the weapon *in the uses for which it was designed* or, as some authorities have put it, its "normal" uses; i.e., the uses for which it was intended'.[20] Second, there is the matter of the indiscriminate use of the weapon, which is covered under targeting law. The principle of distinction or discrimination requires that 'a combatant, using reasonable judgment in the circumstances, distinguish between combatants and civilians, as well as between military and civilian objects'.[21] The following section will turn to the issues that the legal review of weapons poses for automated cyber operations. It will argue that the distinction between nature and use is empty for automated cyber operations and will propose a normative review to prevent the launch of indiscriminate cyber operations.

17 'Paris Call for Trust and Security in Cyberspace', 12 November 2018, https://www.diplomatie.gouv.fr/IMG/pdf/paris_call_cyber_cle443433-1.pdf.

18 GCSC, 'Advancing Cyberstability. Final Report of the Global Commission on the Stability of Cyberspace', November 2019, https://cyberstability.org/report/.

19 Liisi Adamson, 'International Law and International Cyber Norms: A Continuum?', in *Governing Cyberspace: Behaviour, Power and Diplomacy*, eds. Dennis Broeders and Bibi van den Berg (London: Rowman & Littlefield, 2020).

20 Anderson and Waxman, 'Debating Autonomous Weapon Systems, Their Ethics, and Their Regulation under International Law', 1105.

21 Ibid.

3. A NORMATIVE REVIEW FOR CYBER OPERATIONS?

The two-fold normative categorisation detaching indiscriminate 'use' from indiscriminate 'nature' has long been a part of the legal debate on cyber operations.[22] The *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (*Tallinn Manual 2.0*) in fact folds, in its Rule 103, the indiscriminate use and nature dichotomy into its definition of cyber weapons, understood as 'cyber means of warfare that are used, designed, or intended to be used to cause injury to, or death of, persons or damage to, or destruction of, objects'.[23] The definition of cyber weapons is thus embedded into that of cyber attacks (Rule 92) insofar as cyber weapons are intended to execute cyber attacks.[24] In addition, through Rule 105, the *Tallinn Manual 2.0* prohibits cyber weapons that are 'inherently indiscriminate' and can be considered, fundamentally, as 'shots in the dark'. In particular, this rule defines that 'means or methods of cyber warfare are indiscriminate by nature when they cannot be: (a) directed at a specific military objective, or (b) limited in their effects as required by the law of armed conflict and consequently are of a nature to strike military objectives and civilians or civilian objects without distinction'. Separating intentional use from 'natural' capability constitutes, however, a problematic endeavour in the assessment of cyber attacks, particularly when it comes to automated operations. Malware-like and automated cyber attacks propagate and detect unpatched vulnerabilities automatically, and thus their intentionality becomes a question of pure design. In these terms, the *modus operandi* of automated malware defies the very 'nature' vs 'use' dichotomy associated with indiscriminate attacks.

The indiscriminate use of a cyber weapon has also been traditionally defined in relation to the type of harm caused (as evidenced by Rule 103 above). This is also problematic, however, because, putting aside the fact that IHL does not apply below the threshold of armed attack,[25] the rules applying to weaponry tend to govern primarily physical effects of the kind that malware seldom achieves.[26] For example,

22 Herb Lin, 'Cyber Conflict and International Humanitarian Law', *International Review of the Red Cross* 94, no. 886 (Summer 2012): 515–31; Michael N. Schmitt and Sean Watts, 'The Decline of International Humanitarian Law Opinio Juris and the Law of Cyber Warfare', *Texas International Law Journal* 50 (2015): 189.

23 Michael N. Schmitt, ed., *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge: Cambridge University Press, 2017), 452.

24 It must be noted that the manual explicitly rules out "the destruction of data" from its definition of cyber attack, unless the destruction of data leads to physical harm. For an alternative perspective, see: Kubo Mačák, 'Military Objectives 2.0: The Case for Interpreting Computer Data as Objects under International Humanitarian Law', *Israel Law Review* 48, no. 1 (March 2015): 55–80, https://doi.org/10.1017/S0021223714000260.

25 Specifically with reference to the legal obligation to conduct a cyber weapons review, Kudláčková et. al. find that there is no legal obligation for States to conduct a weapons review outside Article 36 of Additional Protocol I, which is not triggered below the threshold of armed conflict. See: I. Kudláčková, D. Wallace, and J. Harašta, 'Cyber Weapons Review in Situations Below the Threshold of Armed Conflict', in *2020 12th International Conference on Cyber Conflict (CyCon)*, vol. 1300, 2020, 97–112, https://doi.org/10.23919/CyCon49761.2020.9131728.

26 See: CCDCOE, 'Scenario 10: Legal Review of Cyber Weapons', Cyber Law Toolkit, n.d., https://cyberlaw.ccdcoe.org/wiki/Scenario_10:_Legal_review_of_cyber_weapons.

when analysing NotPetya in light of this effects-based criterion for indiscriminate use, it becomes apparent that the fake ransomware, which destroyed data, hardly compares to the physical harm caused by a weapon. Above all, this indicates how a harm-based understanding of cyber weapons hinders the protection of networks below the threshold.

States have recently attempted to address this obstacle by 'softening' the harm requirement in their interpretations of how IHL, and in particular Additional Protocol I to the Geneva Conventions,[27] applies to cyberspace. For example, France officially embraced a softer requirement for a cyber operation to rise to the level of armed attack (as defined for the purpose of IHL in Article 49 of Additional Protocol I[28]): for France, it suffices that the cyber weapon disables systems to the point that they are incapacitated 'to provide the service for which they were implemented, whether temporarily or permanently, reversibly or not'.[29] By doing so, the definition of cyber weapons interestingly shifts from harm to effects.

The legal instrument tasked with assessing the conformity of new cyber weapons to IHL standards has been outlined through *Tallinn Manual 2.0*'s Rule 110, which translates the legal review of weapons (as instituted by Article 36 of Additional Protocol I[30])to the context of cyber operations: 'All states are required to ensure that the cyber means of warfare that they acquire or use, comply with the rules of the law of armed conflict that bind them.' While constituting an important tool for the safeguarding of the principles of distinction and discrimination in wartime, a standard legal review of cyber weapons – as prescribed by the *Tallinn Manual 2.0* – cannot be applied to automated cyber operations below the threshold of armed conflict.[31] Additionally, as software remains subject to frequent changes and self-remodulations, it would be impractical to provide a new standard legal review each time software is edited.[32]

With State-sponsored cyber operations primarily occurring in peacetime and with no physical harm, a different normative approach is required to assess them and to

27 The Additional Protocol I to the Geneva Conventions (1977) is a fundamental document for IHL as it reaffirms and modernizes the principles of the original Geneva Conventions (1949), https://ihl-databases.icrc.org/ihl/INTRO/470.

28 Article 49 of the Additional Protocol I states that: '"Attacks" means acts of violence against the adversary, whether in offence or in defence.'

29 French Ministry of the Armies, 'International Law Applied to Operations in Cyberspace', September 2019, https://www.defense.gouv.fr/content/download/567648/9770527/file/international+law+applied+to+operations+in+cyberspace.pdf.

30 Article 36 of the Additional Protocol I states that: 'In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party.'

31 Natalia Jevglevskaja, 'Weapons Review Obligation under Customary International Law', *International Law Studies* 94 (2018): 186–221.

32 Gary Brown and Andrew Metcalf, 'Easier Said than Done: Legal Reviews of Cyber Weapons', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 12 February 2014), 133, https://doi.org/10.2139/ssrn.2400530.

prevent the indiscriminate infliction of damage. This article therefore proposes to take the spirit of the principles of distinction and discrimination outside the stringent legal framework of IHL as a negative norm that: 1. applies to indiscriminate cyber operations below the threshold; 2. regards the effects of cyber operations beyond 'physical harm'; 3. proposes a 'normative' review of cyber operations focusing on nature/design; and thus, 4. creates a normative benchmark for responsible State behaviour, which can be used to hold States accountable when they fail to prevent the viral spread of their cyber operations. With the distinction between indiscriminate use and indiscriminate nature blurring away in the context of automated cyber operations, a normative review of cyber operations should prioritise the assessment of their 'design'. To this end, we use a conceptualisation of a cyber operation that primarily focuses on its nature and thus constitutes 'the combination of a propagation method, exploits, and a payload designed to create destructive physical or digital effects'.[33] Envisioning nature and use as inseparable, the next section will discuss on the basis of two contrasting cyber operations how a normative review can reveal a cyber operation's discriminate or indiscriminate design.

4. SETTING THE STANDARD IN PRACTICE: COMPARING STUXNET AND NOTPETYA

Applying a norm against indiscriminate sub-threshold cyber operations would require States to conduct a review of each operation to ascertain that the design of the operation reflects the attacker's intent to limit its uncontrolled harmful effects (including the destruction of data). This section will demonstrate how the post-incident forensic analysis of an operation, including the reverse-engineering of malware, can allow third parties and victims to determine whether an attacker conformed to the norm. Such analyses are important, as their findings provide a basis on which States can call out unacceptable behaviour and thus set a normative benchmark. Using the examples of NotPetya and Stuxnet, the section will demonstrate how, once the malware had been found 'in the wild', that is, once the malware had spread among real-world computers (as opposed to test systems),[34] interested parties were able to determine whether the operations were indiscriminate in nature from the technical analysis of the malware code.

NotPetya

NotPetya, while masquerading as ransomware, in fact irreversibly encrypted every infected machine's master boot record, thus effectively destroying these computers.[35] As a result of the operation, Maersk, just one of NotPetya's many victims, reported

33 Trey Herr, 'PrEP: A Framework for Malware & Cyber Weapons', *Journal of Information Warfare* 13, no. 1 (2014): 87–106.

34 Trend Micro, 'In-the-Wild - Definition - Trend Micro USA', Trend Micro, accessed 23 November 2020, https://www.trendmicro.com/vinfo/us/security/definition/in-the-wild.

35 Buchanan et al., 'Automating Cyber Attacks', 9.

the loss of 49,000 laptops and 3,500 servers.[36] There are two technical characteristics of NotPetya which reveal that it did not conform to the normative principle of discrimination at the design stages. First, an analysis of what the MITRE ATT&CK Framework[37] terms the 'initial access vector' stage of the operation reveals that the attackers compromised the software update system for the M.E.Doc financial application.[38] In order to engineer this, they had first stolen the credentials of an M.E.Doc administrator to gain control of M.E.Doc's upgrade server to modify the software update so that it would include a 'backdoored' module.[39] M.E.Doc is the most popular accounting software in Ukraine, used widely by any organisation that files taxes or conducts business in the country, including multinational corporations.[40] The attackers' choice of M.E.Doc as the attack vector, therefore, already suggests that the attackers did not pay attention to distinguishing between targets. In addition, when installed by users, the malicious update allowed the attackers to collect email usernames and passwords from organisations that use M.E.Doc and their EDRPOU numbers; these numbers are unique legal entity identifiers given to every organisation that conducts business in Ukraine.[41] The fact that the attackers engineered the malware to collect the numbers is important, as it indicates that they intended for it to spread widely and wanted to identify exactly which organisation was running the backdoored M.E.Doc software.[42] We can therefore conclude from just the analysis of NotPetya's method of delivery that it was not designed to discriminate between systems in its method of delivery.

The analyses of the 'lateral movement' stage, in which the adversary moves from one system to the next within a network, and the 'impact'[43] stage, where the adversary tries to manipulate, interrupt, or destroy systems or data, reveal a second important characteristic: the malware's high level of automation and inability to distinguish between targets before installing and releasing its payload. These stages of the operation indicate that NotPetya's designers did not seek to limit in any way the malware's uncontrolled harmful effects.

36 Rae Ritchie, 'Maersk: Springing Back from a Catastrophic Cyber-Attack | I-CIO', I - Global Intelligence for Digital Leaders, August 2019, https://www.i-cio.com/management/insight/item/maersk-springing-back-from-a-catastrophic-cyber-attack.

37 The framework is a matrix of adversary tactics and techniques based on real-world observations, which in a post-mortem analysis of an operation helps determine the actions an attacker might have taken. See: The MITRE Corporation, 'Matrix: Enterprise | MITRE ATT&CK™', MITRE ATT&CK™, 2018, https://attack.mitre.org/matrices/enterprise/; Chris Brook, 'What Is the MITRE ATT&CK Framework?', Digital Guardian, 23 April 2020, https://digitalguardian.com/blog/what-mitre-attck-framework.

38 Mark Simos, 'Overview of Petya, a Rapid Cyberattack', Microsoft Security, 5 February 2018, https://www.microsoft.com/security/blog/2018/02/05/overview-of-petya-a-rapid-cyberattack/.

39 David Maynor et al., 'The MeDoc Connection', Cisco Talos (blog), 5 July 2017, http://blog.talosintelligence.com/2017/07/the-medoc-connection.html; Anton Cherepanov, 'Analysis of TeleBots' Cunning Backdoor', WeLiveSecurity, 4 July 2017, https://www.welivesecurity.com/2017/07/04/analysis-of-telebots-cunning-backdoor/.

40 Greenberg, *Sandworm*, 179; Maynor et al., 'The MeDoc Connection'.

41 Cherepanov, 'Analysis of TeleBots' Cunning Backdoor'.

42 Ibid.

43 The MITRE Corporation, 'Software: NotPetya | MITRE ATT&CK™', MITRE ATT&CK™, 2018, https://attack.mitre.org/software/S0368/.

In order to propagate across systems, the NotPetya malware used a number of methods.[44] The first, and most effective, was the use of a modified version of Mimikatz, a popular open-source tool used to steal user login credentials from computer memory.[45] Once it had recovered the Windows login credentials from the machine of an infected administrative user, the malware used common Windows management tools to spread itself automatically to other systems on the same network.[46] The second method used by the malware to propagate was through the use of the EternalBlue exploit tool. EternalBlue utilises the CVE-2017-0144 vulnerability in the Server Message Block (SMB) protocol[47] on unpatched Windows systems to allow attackers to remotely infect all the systems on a given network in minutes.[48] By designing the malware so that it used not only EternalBlue but also the modified version of Mimikatz, the attackers ensured that it would self-propagate even to machines that were running an updated version of Windows.[49] NotPetya was therefore designed to behave like an automated worm, spreading via trusted networks rather than the internet, which meant that it bypassed the processes put in place to prevent ransomware attacks.[50] The presence of modified Mimikatz and EternalBlue in the malware code reveals that it was not intended to discriminate between targets, but instead was designed to propagate as widely and as quickly as possible. In fact, coupling automated credential theft and re-use with vulnerability exploitation was what made NotPetya uniquely able to propagate on the widest scale in the history of cyber attacks.[51] Most crucially, however, the malware had no mechanism to distinguish between targets prior to installing its payload: once it had spread to a new host, it automatically scanned other systems for their vulnerability to the SMB exploit in order to release its payload there as well.[52] Therefore, the indiscriminate, automated propagation and installation of malware meant that the destruction wrought by NotPetya had global ramifications.

Stuxnet

Like NotPetya, Stuxnet was also a worm with the capacity to propagate automatically, but Stuxnet serves as a good example of how a technical analysis of an operation can reveal the attackers' intent to limit indiscriminate spread and destruction. In the initial

44 CISA, 'Petya Ransomware', Cybersecurity & Infrastructure Security Agency, 1 July 2017, https://us-cert.cisa.gov/ncas/alerts/TA17-181A.

45 Alexander Chiu, 'New Ransomware Variant "Nyetya" Compromises Systems Worldwide', Cisco Talos (blog), 27 June 2017, http://blog.talosintelligence.com/2017/06/worldwide-ransomware-variant.html.

46 CISA; James Maude, 'NotPetya Ransomware: Attack Analysis | BeyondTrust', BeyondTrust, 20 October 2017, https://www.beyondtrust.com/blog/entry/notpetya-ransomware-attack-analysis; Greenberg, *Sandworm*, 182.

47 The SMBv1 protocol is a network communication protocol that was developed in 1983 to enable computers on a network to share access to files, printers, and ports. See: Carly Burdova, 'What Is EternalBlue and Why Is the MS17-010 Exploit Still Relevant?', Avast, 18 June 2020, https://www.avast.com/c-eternalblue.

48 CISA.

49 Greenberg, *Sandworm*, 182.

50 NCSC, 'Russian Military "Almost Certainly" Responsible for Destructive 2017 Cyber Attack', National Cyber Security Centre, 14 February 2018, https://www.ncsc.gov.uk/news/russian-military-almost-certainly-responsible-destructive-2017-cyber-attack.

51 Simos.

52 CISA.

access stage, in order to deliver the first iteration of the Stuxnet malware into the systems of the Natanz uranium enrichment plant in Iran, which was air-gapped, the perpetrators recruited a mole to physically infect a USB flash drive with the malware, which was then plugged into the centrifuge systems at the plant.[53] Prior to delivering the malware on the USB drive, the mole had visited Natanz a number of times in order to collect detailed information on the configuration of its systems. This allowed the attackers to update the code several times before launching the operation and ensure that the malware would only deliver its payload when it found a very specific configuration of equipment and network conditions (this stage will be elaborated on later).[54] An analysis of the intrusion vector for this first version of Stuxnet reveals that it was designed as a 'precision attack': the malware was injected into only one target network, that of the Natanz facility, and was intended to spread to systems only 'within' that network.[55]

In the second iteration of the operation, to deliver a modified version of the malware, rather than using a mole, the attackers infected the systems of five unwitting external Natanz contractors.[56] It was this change in the malware's delivery, which meant that it eventually spread outside Natanz. Although the malware was designed to only propagate automatically in 'trusted networks', the infection of the contractors' systems meant that the malware spread to the contractors' other customers, most likely through removable drives. It then spread through trusted networks, which are often channelled via the internet, and ultimately ended up infecting over 100,000 computer systems globally.[57] It was at this stage that the malware 'simply went off task'.[58] However, comparing the lateral movement stages of the NotPetya and Stuxnet operations, there is one crucial difference: while the malware spread far and wide in both cases, Stuxnet did not destroy any systems that were not its intended target because it was designed to only deliver its payload to specific types of Simatic programmable logic controller (PLC) devices.[59] Having detected a Simatic PLC, Stuxnet then verified whether it was connected to a specific type of frequency converter running at 807–1,210 Hz, which was the range within which Natanz was known to run its centrifuges.[60] When Stuxnet detected these specific configurations, it released its payload, causing the PLCs to run at different speeds; when it did not, it withheld the payload.[61] Therefore, although

53 Kim Zetter and Huib Modderkolk, 'Revealed: How a Secret Dutch Mole Aided the U.S.-Israeli Stuxnet Cyberattack on Iran', Yahoo News, 2 September 2019, https://news.yahoo.com/revealed-how-a-secret-dutch-mole-aided-the-us-israeli-stuxnet-cyber-attack-on-iran-160026018.html.

54 Ibid.

55 Ibid.

56 Ibid.

57 Ralph Langner, 'To Kill A Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve', *The Langner Group*, November 2013, 18.

58 Kaspersky, 'Stuxnet: Victims Zero', Kaspersky Daily, 18 November 2014, https://www.kaspersky.com/blog/stuxnet-victims-zero/6775/.

59 Nicolas Falliere, Liam O. Murchu, and Eric Chien, 'W32.Stuxnet Dossier', Symantec, November 2010, https://www.wired.com/images_blogs/threatlevel/2010/11/w32_stuxnet_dossier.pdf.

60 Jon R. Lindsay, 'Stuxnet and the Limits of Cyber Warfare', *Security Studies* 22, no. 3 (2013): 383.

61 Lindsay, 383; Michael Lee, 'Stuxnet Infected Chevron, Achieved Its Objectives', ZDNet, 9 November 2012, https://www.zdnet.com/article/stuxnet-infected-chevron-achieved-its-objectives/.

Stuxnet spread in a worm-like fashion, it did not have uncontrolled harmful effects as the malware did not release the payload in systems outside Natanz.[62] For example, Chevron, the energy company, was infected by the Stuxnet malware, but its systems did not sustain any damage.[63] In fact, so precise was Stuxnet's targeting capability that Richard Clarke, a former long-term US counterterrorism chief, commented that it felt like it had been 'written by or governed by a team of Washington lawyers' to limit its collateral damage.[64] We can therefore conclude that because Stuxnet withheld its payload outside Natanz, the spread to other networks outside the Iranian nuclear plant was highly likely to have been unintentional, while the avoidance of indiscriminate harmful effects was fully intentional. Consequently, the analysis of Stuxnet's code reveals design features which indicate that, unlike NotPetya, it complied with the norm of discrimination. Table I compares and summarises the two operations.

TABLE I: SUMMARY OF FINDINGS FROM THE ANALYSIS OF NOTPETYA AND STUXNET MALWARE

	NotPetya	Stuxnet
Initial access vector	Via backdoor implanted in M.E.Doc software update known to be used widely by civilians in Ukraine. This shows that the malware was meant to enter thousands of networks.	Via an external drive inserted directly into a single target network; and via the machines of 5 external contractors known to work at Natanz. This shows the malware was meant to enter only one network.
Lateral movement	Via trusted networks using Mimikatz and EternalBlue. This shows the malware was meant to spread rapidly into every system on the thousands of networks it entered.	Via trusted networks using a number of vulnerabilities including zero days. This shows that the malware was intended to spread, but from the 'impact' stage we can determine that the designers most likely wanted the malware to spread only in the Natanz network.
Impact	Release of malicious payload regardless of environment specifications.	Release of payload only under very specific conditions.

As the two examples illustrate, the ability to reverse-engineer malware once it has been found 'in the wild' provides a basis for judging if in the design of a cyber operation the perpetrator has complied with the norm against indiscriminate operations. In particular, an operation can be judged as indiscriminate if the analysis reveals that the malware contained no mechanism for distinguishing between 'innocent' systems and its intended target prior to installing its payload. Other indications of an operation's lack

62 It is important to note that although the Stuxnet malware did not release its payload in non-target systems, the attackers chose not to delete the malware from non-target systems, despite most likely having the capability to do so. See: Kim Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*, First Edition (New York: Crown Publishers, 2014).

63 Lee, 'Stuxnet Infected Chevron'.

64 Clarke cited in: Ron Rosenbaum, 'Richard Clarke on Who Was Behind the Stuxnet Attack', *Smithsonian Magazine*, April 2012, https://www.smithsonianmag.com/history/richard-clarke-on-who-was-behind-the-stuxnet-attack-160630516/.

of attention to discrimination in the design stages are the malware's infection vector and spreading mechanisms. If the initial access vector targets thousands of networks simultaneously, it raises the likelihood that the operation will be indiscriminate. In terms of propagation, as the analysis of Stuxnet showed, the incorporation of propagation mechanisms into the malware in itself does not necessarily indicate the attacker's lack of intent to limit the operation. Instead, propagation coupled with the malware's inability to distinguish between systems in delivering its payload is what betrays the attackers' inattention to discrimination.[65]

5. CONCLUSION

This paper has argued that the legal principles of distinction and discrimination provide inspiration for a new norm that addresses automated and indiscriminate cyber attacks below the threshold of the use of force. It showed, first, that States have started to articulate a demand for such norms, as they are increasingly concerned about indiscriminate, automated cyber operations. Second, the paper argued that to ensure compliance with the norm in the context of automated cyber attacks, the IHL distinction between the nature of the capability and the use of the capability becomes meaningless, shifting the emphasis to the notion of an 'indiscriminate cyber operation'. States should focus on reviewing the design of cyber operations to ensure that they avoid indiscriminate damage. In other words, if an automated operation is in its 'nature' designed to avoid indiscriminate damage, then its 'use' will be a direct reflection of that design. Third, the paper showed that reverse engineering malware after it has been found 'in the wild', which is routinely done in the aftermath of an operation to establish the attack's source, also allows for a determination of whether a cyber operation's designer sought to limit the harmful effects of the malware to non-target systems. Such forensic analyses are important as they provide a basis upon which States may determine if the attackers conformed with the norm and thus allow them to call out unacceptable behaviour (as part of their public attribution statements, for example) and set a normative benchmark.

65 It is important to note that for the purposes of assessing compliance with the norm, it is irrelevant whether an operation was intentionally indiscriminate or indiscriminate due to coding errors or unforeseen interactions between systems. Indiscriminate spread due to negligence constitutes a breach of the norm.