CYBER WARFARE AND AUTONOMOUS SELF-DEFENCE

Francis Grimal[*] and *Jae Sundaram*[**]

> The enemy will be different.... No longer will it be the simple terrorist armed with an AK-47 or the Semtex bomb (although he will still be around); the new threat will be groups who will bond in cyber space and attack using the new weapons of war: viruses, bugs, worms and logic bombs—James Adams, *The Next World War: Computers Are the Weapons and the Front Line is Everywhere* (1998).

Abstract

The last two decades have witnessed increased activity by states within the 'fifth dimension' (cyberspace) to conduct both civilian and military operations. It is now over two decades since Arquilla and Ronfeldt warned about the advent of cyber warfare in the foreseeable future, and cyberspace has now become a primary battlefield. Prevailing at the forefront of academic scrutiny within the *jus ad bellum* context is the extent to which cyber operations fall within the paradigm of Article 2(4) of the United Nations Charter. A traditional and restrictive interpretation of the cornerstone prohibition contained in Article 2(4) would conclude that the type of force (either threatened or actual) would need to be military / kinetic, thus potentially excluding the possibility of cyber activities. Naturally, some states would contest that it is the consequence *suffered* rather than the modality of attack.

In turn, this raises issues as to whether or not the injury suffered by a state subjected to a cyber attack would be sufficient to invoke its inherent right of self-defence. The scope of this article is to consider the natural technological trajectory of self-defence in cyber operations by examining the very real possibility that computer networks may be enabled to eventually seek to automatically defend themselves against more aggressive cyber intrusions—'automated cyber self-defence'. This would therefore necessitate an examination of the way and extent to which such actions would fall within the existing framework regulating a defensive response. More controversially, the article will also assert that the temporal parameters of self-defence in response to a cyber attack may need re-calibration—issues of detection (particularly against dormant malware etc.) and attribution would prevent a state from responding in a more conventional timeframe. Would self-defence therefore be permissible or indeed desirable several months after an attack has occurred if it is only then attribution becomes clear?

Key Words: Cyber-Threat; Cyber-Defence; Necessity; Proportionality;

[*] University of Buckingham, UK.
[**] University of Buckingham, UK.
The authors would like to express their sincerest thanks to both Professor James A. Green (University of Reading, UK), Dr Duncan Hodges (Cranfield University, UK), and Professor Dr Tom Ruys (Ghent University, Belgium) for their kind and invaluable comments during the earlier stages of drafting this Article.

Attribution; Automated Cyber Self-Defence

I. INTRODUCTION

A cyber attack carried out against a military establishment is capable of devastating a state's defences, or indeed severely limiting a state's abilities to develop its military capabilities.[1] Equally, cyber operations against key institutional infrastructures have the ability to paralyze the day-to-day operations of a state. The increasing number of cyber attacks on state institutions[2] and military establishments have witnessed modern states seeking to build a robust cyber-defence mechanism to thwart potential attacks.[3] Nevertheless, the nature and unique attributes of networked technology require additional work to clarify how the laws may apply to cyber self-defence and automated self-defence, when dealing with cyber attacks.[4] Commentators such as Maogoto and Nguyen note that the 'definitional boundaries remain blurred' in the realm of cyber attacks, as international law provides no direct guidance as to when a cyber attack could rise to the level of an armed attack.[5] This inevitably raises the question on the legality of any measures taken in self-defence, including automated self-defence, and anticipatory measures that may come to

[1] The Stuxnet virus was allegedly used to thwart the attempts of Iranian government in pursuing its nuclear capabilities. The Stuxnet worm's resemblance to legitimate software, such as digital certificates, while using a self-launching programme, allowed its rapid, unobstructed distribution to make it appear trustworthy and later take control of the centrifuges in Natanz, Iran. See Jon R. Lindsay, 'Stuxnet and the Limits of Cyber Warfare' (2013) 22(3) *Security Studies* 365-404. Also to be mentioned is the cyber attack on Georgia carried out in August 2008, just before Russian military forces entered its borders, which affected government websites, news outlets and even Georgia's largest bank. This attack closely resembled the earlier attacks carried out in Estonia in 2007. See Andrew M. Colarik and Lech Janczewski D.Eng, 'Establishing Cyber Warfare Doctrine,' (2012) 5(1) *Journal of Strategic Security* 31-48.

[2] In February 2016, a cyber attack was carried out on the national bank of Bangladesh (Bangladesh Bank), with the attackers managing to syphon out over US$80 million. See Serajul Quadir, 'Malware Suspected in Bangladesh Bank Heist: Officials' *Reuters News*, 11 March 2016, <http://www.reuters.com/article/us-usa-fed-bangladesh-malware-idUSKCN0WD1EV> (accessed 12 December 2016).

[3] See generally, Jeffrey Carr, *Inside Cyber Warfare* (O'Reilly Media, 2012) 161-77, where the author while examining the military doctrines of cyber warfare in the Russian Federation, People's Republic of China (PRC) and the USA, notes that over 120 nations are currently engaged in developing cyber warfare capabilities.

[4] *White House Cyber Strategy* (2011) 9; See Michael N. Schmitt (ed.), *Tallinn Manual on the International Law Applicable to Cyber Warfare* (Cambridge, 2013) 3, where the author notes that the experts were unanimous in their estimation that both the *jus ad bellum* and *jus in bello* apply to cyber operations. See also Reese Nguyen, 'Navigating *Jus Ad Bellum* in the Age of Cyber Warfare' (2013) *California Law Review* 1079-1130. The author opines that despite cyberspace being the new battlefield for nation-states in conflict, *jus ad bellum* provides little guidance about the legality of a cyber attack or when an attack could be viewed as an act of war justifying an armed response.

[5] See Jackson Maogoto, Technology and the Law on the Use of Force **(**Routledge Press, 2015), 12; (Nguyen n 4). See also statement of Keith Alexander, Nominee, Commander, U.S. Cyber Command, 'There is no international consensus on a precise definition of a use of force, in or out of cyberspace. Consequently, individual nations may assert different definitions, and may apply different thresholds for what constitutes a use of force.' (2010) *Armed Services, 110th Servs., 111th Cong*.

be taken to counter any potential cyber attacks.

The structure of this article is set out as follows: Part II of this article will place this discussion within context by examining some of the key and most significant cyber operations to date—the main purpose of which, is to highlight the perennial difficulty of attribution when it comes to cyber warfare. Consequently, the Overview in Part II will avoid any ancillary discussion regarding IHL as that is outside the scope of this article. Meanwhile, Part III of the article will set out the groundwork and necessary framework within which, the concept of automated cyber self-defence, which, we wish to advance in Part IV can be grounded. Part IV of the article examines the unique key thesis of this piece, which, we wish to assert—namely, that the natural trajectory when it comes to cyber attacks is that the Computer Network Infrastructure (CNI) will seek to defend itself automatically against an attack and *may* even seek to anticipate / intercept such an attack. Within the concluding observations, the authors will present the view that the analysis undertaken in Part IV while admittedly, is perhaps more of a theoretical view/exercise at this stage, but the very real possibility of automated responses is not far from reality.

II. AN OVERVIEW OF RECENT CYBER OPERATIONS

The following section (as noted in the introduction) will provide an overview of recent cyber operations/incursions and intrusions but more importantly, familiarise the reader with the applicable 'technology' so as to inform regarding the latter parts of this article. Consequently, the lawfulness or not of such attacks/operations will not be scrutinised—such a task has already been undertaken in quite microscopic detail by others.[6] Rather, the overview will highlight the major difficulties in terms of identifying the attacker and taking appropriate defensive 'action'.

A. Estonia and the Tallinn Data Siege

One of the most popular and earliest methods of cyber attacks to emerge from the 1990s, was the Distributed Denial of Service (DDoS) attack, whereby cyber attackers overwhelm servers (and bandwidth) by bombarding them with unusually heavy bursts of data, or traffic.[7] This is achieved through the use of a network of compromised zombie computers,[8] in which, the owners of

[6] See, for example, James A Green, 'The Regulation of Cyber Warfare under the *Jus ad Bellum*' in James A Green (ed.), *Cyber Warfare: A Multidisciplinary Analysis* (Routledge, 2015), 96; Marco Roscini, *Cyber Operations and the Use of Force in International Law* (Oxford University Press, 2014); and Matthew C Waxman, 'Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)' (2011) 36 *Yale Journal of International Law* 421; Oona Hathaway, Rebecca Crootof, Philip Levitz, Haley Nix, Aileen Nowlan, William Perdue and Julia Spiegel 'The Law of Cyber-Attack' (2012) 100 *California Law Review* 817.

[7] See Susan W Brenner, *Cyberthreats: The Emerging Fault Lines of the Nation State* (Oxford, 2009) 1. The goal of a DDoS attack is to make the use of the network impossible for users, both internal and external. See Jose Nazario, 'Politically Motivated Denial of Service Attacks' in Christian Czosseck and Kenneth Geers (eds), *The Virtual Battlefield: Perspectives on Cyberwarfare* (IOS Press, 2009) 163-164.

[8] Unsuspecting computers are taken over by software that subtly and invisibly infiltrates a

infrastructure will be oblivious to the fact that their equipment has been completely compromised, and taken over by bot networks (botnets).[9] The botnets then become part of a network of slave computers.[10] DDoS attacks are viewed by some as an extension of politics in the 21st century, as increasing number of attacks are carried out by both state and non-state actors on other states, groups, or political factions.[11] It is to be noted that conclusive attribution of a DDoS attack can be extremely difficult,[12] as attacks can stem from multiple sources, and highly complicated to trace.[13] In most cyber attacks involving the use of malware, some form of backdoor payload is introduced, which can then be

---

computer. They are also referred to as 'bots' and are capable of taking over any computer, private or something used by an institution. See Nichollas Ianelli and Aaron Hackworth, 'Botnets as a Vehicle for Online Crime,' *CERT Coordination Center*, 1 December 2005, <https://resources.sei.cmu.edu/asset_files/WhitePaper/2005_019_001_51249.pdf> (accessed 12 December 2016). See also Felix Leder, Tillmann Werner and Peter Martini 'Proactive Botnet Countermeasures: An Offensive Approach' in Christian Czosseck and Kenneth Geers (eds), *The Virtual Battlefield: Perspectives on Cyberwarfare* (IOS Press, 2009) 211-225. The authors describe botnet as an alliance of interconnected computers infected with malicious software (a bot), which can be commanded by an operator and can typically be advised to harvest information such as license keys or banking data on compromised machines, or even launch DDoS attacks against a chosen target.

[9] Botnets would typically consist of Microsoft Windows machines belonging to small business or home-computers that are ill-secured to viruses. See Brian Krebs, 'Bringing Botnets out of the Shadows,' *The Washington Post*, 21 March 2006, <http://www.washingtonpost.com/wp-dyn/content/article/2006/03/21/AR2006032100279.html> (accessed 12 December 2016).

[10] Brenner (n 7). See also 'The Botnet Trackers,' *The Washington Post*, 16 February 2006, <http://www.washingtonpost.com/wp-dyn/content/article/2006/02/16/AR2006021601388.html> (accessed 12 December 2016).

[11] Nazario (n 7). Toolkits developed in the 1990s to carry out DDoS attacks were quickly adapted for political targets. One of the earliest events include attacks on NATO computers in the former Yugoslavia in the late 1990's, and the attacks carried out by Chinese hackers on US military sites in response to the bombing of the Chinese embassy in the former Yugoslavia by US fighter jets. In more recent times it has been reported that the US presidential election campaign groups of Hillary Clinton have suffered several data breaches, most probably carried out by Russian hackers. See Mark Hosenball, Joseph Menn and John Walcott, 'Clinton Campaign Also Hacked in Attacks on Democrats,' *Reuters News*, 29 July 2016, <http://www.reuters.com/article/us-usa-cyber-democrats-investigation-exc-idUSKCN1092HK> (accessed 12 December 2016). It is now strongly believed that Russia tried to meddle in the 2016 US Presidential elections by hacking into the Democratic National Committee's servers and Senator Hilary Clinton's emails, and later releasing them to Wikileaks for publication. See, Eugene Kiely, 'Trump, Russia and the US Election,' *The Wire*, 8 December 2016, <http://www.factcheck.org/2016/12/trump-russia-u-s-election/> (accessed 12 December 2016); Rebecca Shabad, 'Obama, GOP Senators Call for Probe to Examine Russia's Meddling in US Election,' *CBS News*, 9 December 2016, <http://www.cbsnews.com/news/obama-gop-senators-call-for-probe-to-examine-russias-meddling-in-u-s-election/> (accessed 12 December 2016).

[12] Messages sent as part of a DDoS attack can enter the internet from any location. This renders it difficult to find the source of the attack by ISPs, and the source addresses assigned to packets as part of a packet flood can be falsified. See Paulo Shakarian, Jana Shakarian and Andrew Ruef (eds.) *Introduction to Cyber-warfare: A Multidisciplinary Approach* (Newnes, 2013).

[13] Two complicating factors in the DDoS attack are, that the attacking computers will be numerous and will change over time, and also, the source addresses (IP addresses) of the attacking computers could be forged. See Subramani Rao & Sridhar Rao, 'Denial of Service Attacks and Mitigation Techniques: Real Time Implementation with Detailed Analysis,' *This paper is from the SANS Institute Reading Room site*, 11 September 2011, <https://uk.sans.org/reading-room/whitepapers/detection/denial-service-attacks-mitigation-techniques-real-time-implementation-detailed-analysi-33764> (accessed 12 December 2016).

used by the attacker to gain access to the infected computer at a later point in time,[14] which again complicates matters further making attribution an extremely difficult task to achieve particularly. This is probably one of the key considerations that re-surfaces in latter discussions in Part IV.

Estonia, a European Union and also a NATO Member State,[15] came under a series of sustained digital attacks on 26 April 2007. The attacks continued for over two weeks, covering various components of Estonia's infrastructure, including state media and the financial sector.[16] Needless to say the cyber attacks were in the form of DDoS attacks, and an investigation of the siege revealed that an estimated 1 million zombie computers were involved in carrying out the attacks, which was unprecedented at that time.[17] By 29 April 2007, a flood of data had shut down the Estonian Parliament, the website of the prime minister, various ministers, and the government name and email servers, besides slowing down transactions of key financial institutions within the state.[18] On 9 May (Victory Day), the attacks peaked, with attacks lasting for 10 hours each, with a peak bandwidth utilization of 95bps.[19]

What the cyber attacks clearly revealed, though was that NATO lacked 'both a coherent cyber doctrine and a comprehensive cyber strategy,'[20] which effectively forced the alliance to reconsider its position and strategy to tackle the growing cyber threat to its member states.[21]

B. Georgia Comes Under Sustained DDoS Attacks

The data siege on Tallinn was just the beginning, as Georgia was soon to experience a tsunami of cyber attacks for a sustained period of time. In July 2008, the DDoS attacks were carried out on Georgian President Mikheil

---

[14] Heather A Harrison Dinniss, *Cyber Warfare and the Laws of War* (Cambridge University Press, 2012) 89-90. The backdoor is a code which opens an undocumented access point to the compromised computer or system by bypassing the authentication and security protocols and allowing the computer for remote access. The malware then is used to recruit the unprotected computers to a network of compromised computers called botnets, which are then in turn used to carry out DDoS attacks.

[15] Estonia is one of the smallest NATO member states, where almost all personal activities such as banking to education is carried out online, which was also a reason for being targeted for cyber attacks. See Vincent Joubert, 'Five Years After Estonia's Cyber Attacks: Lessons Learned for NATO?' *NATO Defense College, Research Division* (2012) No 76.

[16] Mark Lander and John Markoff, 'Digital Fears Emerge after Data Siege in Estonia,' *The New York Times*, 29 May 2007, <http://www.nytimes.com/2007/05/29/technology/29estonia.html?_r=0> (accessed 12 December 2016).

[17] Brenner (n 7). See also Lander and Markoff (n 16) where the authors note that Dutch authorities reportedly encountered 1.5 million botnets some years ago.

[18] Nazario (n 7) 166. The attackers used multiple attack methods, including the use of Russian language forums and blogs to spread tools (such as ping flood scripts). To coordinate their efforts, the attackers also recruited botnets into the effort to fire them at the same time.

[19] Most of the attacks measured in ATLAS died out after Victory Day, although reports from first-hand accounts within Estonia indicate that they continued for several weeks. See Jose Nazario, 'Estonian DDoS Attacks – A summary to date,' *Arbor Networks*, 17 May 2007, <https://www.arbornetworks.com/blog/asert/estonian-ddos-attacks-a-summary-to-date/> (accessed 12 December 2016).

[20] Rex B. Hughes, 'NATO and Cyber Defence' (2009) 33 *Atlantisch Perspectief*.

[21] Joubert (n 15).

Saakashvili's website, with a much more substantial wave of attacks to come on 8 August.[22] In early August, Georgia and Russia exchanged fire and Russian tanks entered Georgian territory, which was instantly followed by a large scale DDoS attacks on the key Georgian sites,[23] including the websites of the president, various ministries, news agencies, and others.[24]

Arbor Peakflow and other traffic monitors on the internet recorded a substantially larger peak size than the attacks carried out in Estonia in 2007.[25] Attempts to load the webpage of the President of Georgia during the attacks, from a number of North American vantage points, were not successful.[26] Analysts also noted that the DDoS attacks appeared to coincide with the Russian troops' movements into South Ossetia, which was in response to Georgian military operations launched a day earlier in the region.[27]

The most important observation to be made here, besides the strong political undertone of the cyber attacks, is the fact that for the first time in over a decade a military conflict and a cyber conflict coincided.[28] The cyber attacks carried out on Georgia were relatively unsophisticated DDoS attacks, and website defacement, but carried out in a very sophisticated manner[29] to achieve maximum results. The primary purpose of the cyber attacks was to lend support to the Russian military operations and to that end the cyber attacks were effective, as it also successfully impeded the Georgian government to deal with the Russian invasion by interfering with communications between the government and the public, besides disrupting the payment mechanisms of the financial institutions.[30] There is no doubt that the attacks were well coordinated between the Russian military campaign and the cyber attackers, as immediately upon the Russian troops establishing their positions within Georgia, the cyber attack list was expanded to include many more government websites.[31]

C. Stuxnet: The Landmark Cyber Attack

[22] On 19 July 2008, an internet security firm reported DDoS attacks on the websites in Georgia. See Stephen W. Korns and Joshua E. Kaestenberg 'Georgia's Cyber Left Hook' (2009) 38 (4) *Parameters* 60-76.

[23] Many of the cyber attacks were so close in time to the corresponding military operations, which leads one to conclude that there had to be close cooperation between people in Russian military and the civilian cyber attackers. The organizers of the cyber attacks should have had advance notice of the Russian military intentions. See 'Overview by the US-CCU of the Cyber Campaign Against Georgia in August 2008' (2009) *A US-CCU Special Report*.

[24] Nazario (n 7) 167.

[25] The peak bandwidth recorded during the attacks was over 800 Mbps, and the attacks were much more intense. See Nazario (n 7) 167.

[26] Steven Adair, 'The Website for the President of Georgia Under Attack - Politically Motivated?' *Shadowserver Foundation Calendar*, 20 July 2008, <http://www.shadowserver.org/wiki/pmwiki.php/Calendar/20080720> (accessed 12 December 2016).

[27] Korns and Kaestenberg (n 22).

[28] Nazario (n 7) 167. The author also notes a more recent Israel-Palestinian conflict where military action coincided with cyber attacks.

[29] US-CCU Special Report (n 23).

[30] Ibid.

[31] Ibid.

Some of the difficulties in determining the lawfulness of actions pursued in cyberspace applying the existing international legal framework were well illustrated in the incident involving the use of Stuxnet virus in 2010 to carry out a cyber attack.[32] If the cyber attacks on Estonia and Georgia were carried with the help of botnets to disrupt state functions and to put pressure on the governments concerned, the Stuxnet virus attack was carried out with utmost precision to paralyse (and very nearly brought to a halt) a state sponsored nuclear programme. Stuxnet is probably the first computer virus known to be capable of targeting and destroying industrial systems such as nuclear facilities and power grids.[33] The Stuxnet worm, a self-replicating computer virus, was used to target the computers used in Iran's nuclear facility to take control of the centrifuges at Natanz, which resulted in the centrifuges spinning out of control and self-destruct.[34]

The worm was designed to sabotage the nuclear programme that Iran was promoting at that time by targeting the industrial control systems (ICSs) of the nuclear facility,[35] and eventually go on to disrupt the nuclear enrichment programme. The seeds for the attack were sown even around 2008, when the worm was first infected networks around the world, although causing no grate damage to most systems infected.[36] Initially it was assumed the attacks on the nuclear facility had not been successful, but in the autumn of 2010, reports spread quickly about Iran's uranium enriching capabilities becoming

---

[32] See Thomas M. Chen, 'Stuxnet, the Real Start of Cyber Warfare? [Editor's Note]' (2010) 24 (6) *IEEE Network* 2-3; Nguyen (n 4).

[33] Jonathan Fildes, 'Stuxnet Worm 'Targeted High-Value Iranian Assets,' *BBC News*, 23 September 2010, <http://www.bbc.co.uk/news/technology-11388018> (12 December 2016). See also Chen (n 32). The author notes that Stuxnet, which is highly selective of its targets, looks for a particular 'programmable logic controller' (PLC) in vulnerable computers, and appears to aim directly at controlling physical machinery.

[34] Thomas N Chen, 'Cyberterrorism After Stuxnet' (2014) *Army War College Carlisle Barracks PA Strategic Studies Institute*; Nguyen (n 4). Because of the 'cyber attack' Iran's uranium enrichment operations halted, resulting in an estimated several years of delay in the country's nuclear arms development program. See Lindsay (n 1). See also Andrew Colarik and Lech Janczewski, 'Establishing Cyber Warfare Doctrine,' (2012) 5(1) *Journal of Strategic Security* 31-48; Aleksandr Matrosov, Eugene Rodionov, David Harley, and Juraj Malcho, 'Stuxnet Under the Microscope,' *ESET LLC, Revision 1.31* (2011) <https://www.esetnod32.ru/company/viruslab/analytics/doc/Stuxnet_Under_the_Microscope.pdf> (accessed 12 December 2016); Paul Mueller and Babak Yadegari, 'The Stuxnet Worm,' *Département des sciences de l'informatique, Université de l'Arizona*, 2012, <https://www.cs.arizona.edu/~collberg/Teaching/466-566/2014/Resources/presentations/2012/topic9-final/report.pdf> (accessed 12 December 2016).

[35] Sean Collins and Stephen McCombie, 'Stuxnet: The Emergence of a New Cyber Weapon and Its Implications' (2012) 7(1) *Journal of Policing, Intelligence & Counter Terrorism* 80-91. The authors also opine that nation states, terrorist groups, hacktivists and cyber criminals to achieve their own goals could use future versions of the virus, and that Stuxnet has started a new arms race, creating serious implications for the security of critical infrastructure worldwide. See also Chen (32). Stuxnet did raise the eyebrows of security researchers for three reasons, namely, its choice of target, level of sophistication, and implications for future malware.

[36] Hathaway, Crootof, Levitz, Nix, Nowlan, Perdue Spiegel (n 6). See also Chen (n 34). Although the primary target was the Bushehr nuclear plant in Iran, the virus infected an estimated 50,000 to 100,000 computers across Iran, India, Indonesia and Pakistan.

diminished.[37] This incident went on to demonstrate how a cleverly carried out cyber attack with surgical precision[38] was capable of causing more harm and damage than a mighty airstrike.

The destructive potential of a carefully carried out cyber attack was well captured in the above incident, as it paralysed the Iranian nuclear facility without coming under a traditional kinetic attack. It left no doubts in the minds of the detractors/analysts that a cyber attack carried out with far less manpower than an aircraft was fully capable of causing more harm and destruction than an army could potentially cause.[39] The Stuxnet virus was crafted to deliver a payload to a specific high-value target clearly designed to bring about real-world damage of ICSs.[40]

The advent of Stuxnet besides revealing the sophistication required for a 'weaponized' malware,[41] has also challenged the popular assumptions prevalent at that time, *i.e.*, that network defences will protect facilities from vulnerabilities in software applications.[42] In Chen's view, Stuxnet now has the attention of the world by promoting an arms race to develop both offensive and defensive cyber capabilities among nations and the underground.[43] It will also not be an exaggeration to state that the incident involving Stuxnet came close to causing a serious human catastrophe, as the cyber attack was carried out on a high value target, *viz.*, nuclear enrichment facility.[44]

---

[37] Chen (n 32). See also 'The Stuxnet Worm: A Cyber-Missile Aimed at Iran?' *The Economist, Babbage Blog*, 24 September 2010, <http://www.economist.com/blogs/babbage/2010/09/stuxnet_worm> (accessed 12 December 2016).

[38] Mark Clayton, 'Stuxnet Malware is 'Weapon' Out to Destroy…Iran's Bushehr Nuclear Plant?,' (2010) 21 *Christian Science Monitor*. The author compares Stuxnet to 'a precision, military-grade cyber missile'.

[39] Chen (n 34) 9. The author estimates the cost of creating the Stuxnet virus to run into millions of dollars, with the use of very substantial resources. The author also opines that Stuxnet has implications for the cost-benefit weights of potential future attacks.

[40] Chen (n 34) 8-9. See also Lindsay (n 1). The attack did not permanently derail Iran's nuclear program, as enrichment recovered within a year, leading to concerns in 2012 that Israel or the US might launch airstrikes to address the problem.

[41] Chen (n 34) 5.

[42] Stuxnet changed a theoretical hypothesis into reality, and there is more likely to be a long-term affect than a short-term one. See Chen (n 34) 9.

[43] (Chen 34) 9. See also Lindsay (n 1), where the author observes that the many now view Stuxnet as the harbinger of even more devastating attacks to come, and that even weaker states and political actors will be encouraged to acquire cyber capabilities, posing a threat to advanced industrial nations.

[44] See Con Coughlin, 'Stuxnet Virus Attack: Russia Warns of 'Iranian Chernobyl" *The Telegraph*, 16 January 2011, <http://www.telegraph.co.uk/news/worldnews/europe/russia/8262853/Stuxnet-virus-attack-Russia-warns-of-Iranian-Chernobyl.html> (accessed 12 December 2016). Dmitry Rogozin, the Russian Ambassador to NATO remarked 'these mines could lead to a new Chernobyl', meaning if the attack had gone wrong it would have led to a nuclear disaster, similar to the Chernobyl nuclear incident of 1986. See David Brunnstrom 'Russia Says Stuxnet could Have caused New Chernobyl," *Reuters News*, 26 January 2011, <http://www.reuters.com/article/us-iran-nuclear-russia-idUSTRE70P6WS20110126> (accessed 12 December 2016). Some have taken the position that the Stuxnet virus was limited in its destruction, and did not cause any fatalities. See also Andrew Futter, 'Hacking the Bomb: Nuclear Weapons in Cyber Age' *International Studies Annual Conference*, February 2015, <http://www2.le.ac.uk/departments/politics/people/afutter/copy_of_AFutterHackingtheBombISAPaper2015.pdf> (accessed 12 December 2016).

Stuxnet has the distinction of being the only historical case available for scrutiny.[45] Whether the Stuxnet cyber attack on the Iranian nuclear facility was in self-defence and justifiable under IHL is highly debatable, as no single state has come forward to assume responsibility for the attack,[46] or has Iran for that matter accused anyone of having disrupted its nuclear enrichment programme,[47] using cyber tactics.[48] Again, the issue surrounding lawfulness of such a cyber attack on a nuclear facility is outside the remit of this article, as others have spent considerable care and attention on examining this particular point.[49] Rather, the more important questions for our current discussion are how a state can justify any self-defence measures taken in cyberspace to prevent a potential cyber strike, and or actively 'anticipate' a cyber strike and counter the same for that matter.

On the flipside, one can also raise the questions of how a state could successfully defend against such strikes, and how can it carry out anticipatory cyber strikes[50] on other states (or non-state actors) to thwart any such cyber attacks which it may come to perceive as being launched. This raises further questions, *i.e.* whether the above defensive strikes and anticipatory strikes can be in response to cyber threats, and in response to non-cyber but military threats.[51] Defensive cyber strikes have the risk of striking the wrong targets and

---

[45] Lindsay (n 1).

[46] Although no official announcement was made, it is widely believed that the cyber attack was carried out through a joint US-Israeli component of a broader US cyber campaign against Iran, code-named 'Olympic Games'. See David E. Sanger, 'Obama Order Sped Up Wave of Cyberattacks Against Iran,' *The New York Times*, 1 June 2012, <http://www.nytimes.com/2012/06/01/world/middleeast/obama-ordered-wave-of-cyberattacks-against-iran.html?_r=0> (accessed 12 December 2016). See also generally Kim Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon* (Crown Publishers, 2014).

[47] Such an admission would have inevitably put Iran in a spot, as it had all along maintained that its nuclear enrichment programme was for civilian purposes and denied that it was capable of producing weapons grade plutonium, and an admission on its true purport would have put the State in trouble with the International Atomic Energy Agency (IAEA). See Chen (n 43), where the author notes that the Iranian officials at the Bushehr nuclear plant categorically denied that the cyber attack caused any damage to their main systems at the nuclear facility. Though the officials appeared to admit to some staff PCs being infected, the two-month delay in bringing the reactor in line was clearly blamed on a leak in a storage pool for the plant's fuel.

[48] Iranian officials were unsure as to what was causing the centrifuges in its nuclear enrichment facility to stall and crash. In short, they were oblivious to the fact that their facility had come under a powerful cyber attack.

[49] Green (n 6), Roscini (n 6), Waxman (n 6), Hathaway, Crootof, Levitz, Nix, Nowlan, Perdue and Spiegel (n 6), and Dinniss (n 14).

[50] See below discussions in Parts III and IV. See also Chen (n 34) 4. The author argues that the US policies can really address only the opportunities for terrorism (but not motive or means) by strengthening the defenses of critical infrastructures; Eric F. Mejia, 'Act and Actor Attribution in Cyberspace, (2014) *Strategic Studies Quarterly* 114-132, 115. See also *International Strategy for Cyberspace: Prosperity, Security, and Openness in Networked World* (Washington: White House, May 2011) 10, where the right of self-defence related to cyber attacks is stated as follows: 'Right of Self-Defence: Consistent with the United Nations Charter, states have an inherent right to self-defense that may be triggered by certain aggressive acts in cyberspace'. It is abundantly clear that the US has taken an affirmative position on the issue as it has become the prime target of cyber attacks in recent years.

[51] The widely-debated involvement of the US and Israel in the Stuxnet cyber attack makes one conclude that it was in response to a non-cyber threat. See Dinniss (n 14) 112-113. Dinniss is of

could potentially run into rough weather, as attackers can compromise third party computers to use as intermediaries, or channels through anonymizing proxies that hide their Internet protocol (IP) address.[52] Dinniss holds the view that Stuxnet, in sharp contrast to the DDoS attacks on Tallinn, and Georgia, resulted in the destruction of property, and would amount to the use of force, but the scale and effects of the attack do not appear to have sufficient gravity to amount to an armed attack.[53]

D. Iranian Hackers and the Cyber Heist on Bangladesh Bank

In stark contrast to the previous examples, more recent cyber operations by Iranian hackers and a Cyber Heist on a Bangladeshi Bank highlight the fact that such activities may not always necessarily involve a kinetic or forceful result. Consequently, this raises the more important question as to what response might be permissible by way of self-defence to actions falling short of activating a forceful defensive measure. [54]

In December 2015, it was reported that an Iranian hacktivist group had claimed responsibility for a cyber attack, which allegedly allowed it to gain access to the control system to a dam in New York.[55] More worryingly, the hackers claimed that they had breached the security even as early as in 2013,

---

the view that it is unlikely that the UN Security Council would consider it necessary to authorize force against a computer network attack. However, when an ongoing series of attacks cannot be stopped by electronic means, it is within the Security Council's purview to authorise force. Dinniss also refers to UN Security Council Resolution 1929 (2010) aimed at Iran, which did not authorise any measures exceeding the scope of the resolution, and notes that had the measures decided on been taken under Article 42 (with its usual phraseology 'all necessary means'), then the Stuxnet worm would have proved an efficient way to achieve one of the aims of the resolution, *i.e.,* suspension of enrichment-related activities. This broader issue is revisited in general terms within Parts III and IV.

[52] Stuxnet used two stolen digital certificates and multiple zero-day exploits making it impossible to attribute the attack carried out to one single source. Chen (n 34) 4-5. The author also notes that the complete effects of a cyber attack may be concealed. For instance, if stealthy malware has been installed without detection, then attribution is difficult. See also Dinnis (n 14), where the author identifies the practice of IP spoofing which is widely used in cyber attacks using botnets, where the identity of the provider is concealed, making attribution difficult. This part is discussed further in Parts III and IV.

[53] Dinniss (n 14) 81-82. The author notes that the worm may have been responsible for Iran having to replace 1,000 of the 9,000 IR-1 centrifuges at the Natanz nuclear fuel enrichment facility. See also David Albright, Paul Brannan and Christina Walrond, 'Stuxnet Malware and Natanz: Update of ISIS December 22, 2010 Report' (2011) *Institute for Science and International Security* 3.

[54] See Mary Ellen O'Connell, 'Cyber Security without Cyber War' (2012) 17 *Journal of Conflict and Security Law* 187, 204-205.

[55] See Danny Yadran, 'Iranian Hackers Infiltrated New York Dam in 2013,' *The Wall Street Journal*, 20 December 2015, <http://www.wsj.com/articles/iranian-hackers-infiltrated-new-york-dam-in-2013-1450662559> (accessed 12 December 2016). Officials feared that hackers had breached the systems at the Arthur R. Bowman Dam in Oregon, a 245-foot-tall earthen structure which irrigates local farm lands and also prevents flooding in Prineville, Oregon which has a population of 9,200. Eventually, it transpired that the hackers had breached the systems at the Bowman Avenue Dam, which is situated near the village of Rye Brook, New York. The dam in question is a 20-foot-tall concrete slab across Blind Brook, about 5 miles from Long Island Sound.

and were prepared to release such technical information to prove it.[56] The hackers also claimed that they did not go public with their attack as there was a 'state-level warning' not to go public. A report into the incident revealed that the hackers were able to access files (including usernames and passwords) six times between 22 August 2013 and 27 September 2013,[57] proving the claims of the hackers. Though the breach was traceable to an Iranian group, it was not clear if the intrusion was condoned by the Iranian government.

However, after a wait of 3 months, the US indicted seven Iranian hackers associated with the cyber attacks against key industries, including breaking into the computer system at a small dam in Rye, NY., and for attacking a US bank's public websites from late 2011 through May 2013.[58] Interestingly, one of the charges brought against the hackers include coordinated 'distributed denial of services,' or DDoS attacks with a view to crash the commercial sites of 46 US financial institutions.[59] This indictment marks the first instance, where the US has charged state-sponsored individuals with hacking to disrupt the networks of key US industries, and it is highly unlikely that the Iranian government will be willing to send those indicted to the US to face trial.[60] Although the US officials were able to complete the investigation more than a year ago, the indictment was held off so as not to jeopardize the landmark 2015 nuclear deal with Iran and a January prisoner swap.[61]

In February 2016, a cyber attack was carried out on the national bank of Bangladesh (Bangladesh Bank), with the attackers managing to syphon out US$81 million. The original plan of the hackers was to embezzle a sum of US$951 million from its Fed account, which it uses for international settlements, but only a typing error made by the hackers prevented the attempts of moving that big amount from the central bank.[62] The key officials in the country were oblivious of the heist, and the Governor of the Bangladesh's Central Bank quit as it emerged a sum of US$30 million, which was stolen from the Bangladesh Bank, was delivered in cash to a casino operator in Manila, Philippines.[63] While the

---

[56] Ibid. When the Wall Street Journal reported the breach in December 2015, SOBH Cyber Jihad decided to go public for the operation against the Bowman Avenue Dam in Rye Brook, NY.

[57] Ibid. Officials in Rye Brook could stress that the hackers did not ever manipulate the dam over Blind Brook.

[58] Ellen Nakashima & Matt Zapotosky, 'National Security: U.S. Charges Iran-Linked Hackers with Targeting Banks, N.Y. Dam,' *The Washington Post*, 24 March 2016, <https://www.washingtonpost.com/world/national-security/justice-department-to-unseal-indictment-against-hackers-linked-to-iranian-goverment/2016/03/24/9b3797d2-f17b-11e5-a61f-e9c95c06edca_story.html> (accessed 12 December 2016).

[59] Ibid. The affected institutions include Bank of America, the Nasdaq Composite Index, the NY Stock Exchange, Capital One, AT&T and PNC.

[60] Ibid.

[61] Dustin Volz and Jim Finkle, 'US Indicts Iranians for Hacking Dozens of Banks, New York Dam,' *Reuters News*, 24 March 2016, <http://www.reuters.com/article/us-usa-iran-cyber-idUSKCN0WQ1JF> (accessed 12 December 2016).

[62] See Quadir (n 2). See also Serajul Quadir 'How a Hacker's Typo Helped Stop a Billion Dollar Bank Heist' *Reuters News*, 10 March 2016, <http://www.reuters.com/article/us-usa-fed-bangladesh-typo-insight-idUSKCN0WC0TC> (accessed 12 December 2016). There is no doubt that the cyber attack carried out on the Bangladesh Bank was not politically motivated but financially motivated.

[63] Serajul Quadir and Karen Lema, 'Man in Manila Gets $30 Million Cash from Cyber Heist; Bangladesh Central Bank Governor Quits,' *Reuters News*, 15 March 2016,

FBI is looking for evidence in the US and beyond to determine who was behind the daring cyber heist,[64] the Bangladesh bank is weighing the options of bringing a court action against the NY Fed over the bank heist carried out through a cyber hack.[65]

It is highly debatable if cyber attacks can be classified as use of force, as it will have to be studied on a case by case basis and any outcome will be largely predicated on any destruction of property, economic consequences brought about by such attacks, the players involved, the value and strategic importance of the targets, and the surrounding political and other circumstances prevailing at the time the attack is carried out. All the above factors play a major rule, and any uncertainty can make attribution an extremely difficult task.

III. LEGAL PARAMTERS

A. *Overview*

In order to evaluate autonomous cyber defensive actions taken by CNIs (contained in Part IV) at a theoretical/conceptual level, this present section will examine the applicable legal framework within which, such a discussion must be grounded. The first caveat with any such exploration is that this section will avoid placing too great an emphasis on overly deconstructing Article 2(4) of the United Nations Charter in the context of cyber operations. This area of the law has been extensively examined and re-examined in recent times.[66]

However, the more important 'discussion' for the purview of this article lies in considering a state's potential responses under Article 51 UN Charter. The law governing a state's inherent right to self-defence traditionally, and to this day, continues to attract considerable forensic analysis.[67] This article will avoid revisiting age-old debates, and instead focus within the specific parameters and remit of the article as set out in the abstract.

Nevertheless, no discussion on cyber operations can completely avoid highlighting the perennial consideration as to whether Article 2(4) fully, or indeed, partly captures the subtleties of a cyber attack. Section B will therefore consider the applicable legal parameters surrounding Article 2(4), while Section

---

<http://www.reuters.com/article/us-usa-fed-bangladesh-governor-idUSKCN0WH0JF> (accessed 12 December 2016).

[64] Abhirup Roy and Nate Raymond, 'FBI Probes Bangladesh Bank Account Cyber-Theft: WSJ' *Reuters News*, 18 March 2016, <http://www.reuters.com/article/us-usa-fed-bangladesh-idUSKCN0WK25L> (accessed 12 December 2016).

[65] Serajul Quadir, 'Bangladesh Bank Weighs Lawsuit Against NY Fed Over Hack,' *Reuters News*, 22 March 2016, <http://www.reuters.com/article/us-usa-fed-bangladesh-idUSKCN0WO2JQ> (accessed 12 December 2016).

[66] See, for example, Green (n 6); Hathaway, Crootof, Levitz, Nix, Nowlan, Perdue and Spiegel (n 6); Dinnis (n 14) 37-74; Roscini (n 6); and Waxman (n 6).

[67] See, for example, as a minimum, Murray Colin Adler, *The Inherent Right of Self-Defence in International Law* (Springer, 2013); Kinga Tibori Szabó, *Anticipatory Action in Self-Defence: Essence and Limits under International Law* (TMC Asser Press, 2011); Tom Ruys, *'Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice* (Cambridge University Press, 2010); James A Green, *The International Court of Justice and Self–Defence in International Law* (Hart, 2009).

C will set out the necessary framework regarding a state's inherent right of self-defence.

B. Deconstruction of Article 2(4) Pertinent to Cyber Operations

The starting point for any discussion concerning *jus ad bellum* considerations requires underscoring the cardinal prohibition against both the threat or use of force by states contained in Article 2(4) of the UN Charter.[68] As some scholars are quick to caution, an overly liberal use of *jus cogens* categorisation is to be avoided,[69] but overwhelmingly, the academic consensus is that the prohibition contained in Article 2(4) has *jus cogens* status, and accordingly, cannot be derogated from.[70] It is important to underline, that it is *not* Article 2(4) per se that has *jus cogens* status, rather, it is the prohibition contained therein.[71] In parallel to the 'negative' prohibition contained in Article 2(4), runs a positive obligation contained in Article 2(3) UN Charter requiring Member States to settle their disputes peacefully. A holistic and combined effect of Article 2(3), Article 2(4) and Article 2(7) alongside the customary principle of non-intervention, is the general obligation against states to interfere within the sovereign affairs of another state.[72]

As highlighted in the abstract, a traditional and undoubtedly restrictive interpretation of the cardinal prohibition contained in Article 2(4) would conclude that the typology of force (whether threatened or actual) would need to be of a military / kinetic nature thus potentially excluding the possibility of cyber activities.[73] In sharp contrast of course, some states would contest that it is the *consequence* suffered rather than the modality of attack, which, would therefore allow cyber-operations to fall within the ambit of Article 2(4).[74] Clearly, at the time in which, the Charter was being drafted and negotiated the use of such advanced technology was not envisaged. Therefore, as many have highlighted, it

[68] This is another area of the *jus ad bellum* which, remains under forensic scrutiny. See, for example, Olivier Corten, *The Law Against War: The Prohibition on the Use of Force in Contemporary International Law* (Hart, 2010), 50-197; Thomas M Franck, *Recourse to Force: State Action Against Threats and Armed Attacks* (Cambridge University Press, 2002), 11-9; and Nico Schrijver, 'The Ban on the Use of Force in the UN Charter' in Marc Weller (ed), *The Oxford Handbook of the Use of Force in International Law* (Oxford University Press, 2015), 466.

[69] James A. Green, 'Questioning the Peremptory Status of the Prohibition of the Use of Force' (2010) 32 *Michigan Journal of International Law* 215.

[70] Alexander Orakhelashvili, *Peremptory Norms in International Law* (Oxford University Press, 2006).

[71] Green (n 69).

[72] Noting, that Treaties are interpreted according to the 1969 Vienna Convention on the Law of Treaties (VCLT). See also Grigoriĭ Ivanovich Tunkin, *Theory of International Law* (Wildy, Simmonds and Hill, 2003) 141.

[73] For that more traditional interpretation, see Daniel B Silver, 'Computer Network Attack as a Use of Force under Article 2(4)' (2002) 76 *International Law Studies* 73, 80-2; and Tom J Farer, 'Political and Economic Coercion in Contemporary International Law' (1985) 79 *American Journal of International Law* 405.

[74] For a classic interpretation, see Ian Brownlie, International Law and the Use of Force by States (Clarendon Press, 1963), 362.

is up to the international legal community to re-evaluate the legal framework given that the latter has been 'overtaken' by advances in technology.[75]

A point made by one of the authors of this Article in a co-authored Book Review (with Professor James A Green) elsewhere, is that an application of *jus ad bellum* created prior to the creation of modern technology will undoubtedly be problematic,[76] as transposing and applying the existing tapestry to modern threats cannot be the solution.[77] Some commentators however, over-play this trajectory.[78] One could safely posit that while cyberspace is unique, it is not *so* unique so as to make the *jus ad bellum* inapplicable.[79] It should not be forgotten that since its creation, the application of Charter Law has necessitated regular 'updates' to contend with emerging threats.[80] Charter laws have been adapted to threats posed in different dimensions, and cyberspace is yet another dimension within which, its application is sought. And finally, even if one does conclude that certain cyber operations may fall outside the purview of the *jus ad bellum*, it is patently incorrect to assert that these would also fall outside the scope of international law generally. Such operations would still be captured within the more general principles of non-intervention[81] and/or the duty of due diligence.[82] While violations of these international norms do not trigger or allow for forcible actions in self-defence, states may still lawfully have recourse to non-forcible countermeasures.[83]

By way of basic deconstruction of Article 2(4), the ICJ in *Nicaragua* employed a fairly liberal definition of 'force', which captures both direct and indirect use of force.[84] Within the context of cyber operations, and specific to this article, one might also raise an interesting but perhaps niche point as to whether the mere 'threat' of cyber force would also fall within the purview of Article 2(4) in terms of constituting an unlawful threat of force? This issue will be revisited in due course in Part IV particularly, since an autonomous response in self-defence may be 'anticipated' against an imminent threat which, has yet to materialise into a 'concrete' armed attack.

Considerable ink has already been spilt in terms of typology assessment regarding cyber attacks and the way in which, they may be captured (or not)

---

[75] See, for example, Maogoto (n 5.).
[76] Francis Grimal and James A Green, 'Technology and the Law on the Use of Force' (2016) 3(1) *Journal of Use of Force and International Law,* at 177-184.
[77] Ibid.
[78] Ibid.
[79] Ibid.
[80] See Wolff Heintschel von Heinegg, 'Territorial Sovereignty and Neutrality in Cyberspace' (2013) 89 *International Legal Studies* 123, 123-4; and Christopher P M Waters, 'New Hacktivists and the Old Concept of *Levée en Masse*' (2014) 37 *Dalhouse Law Journal* 771, 773-5.
[81] See Russel Buchan, 'Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions' (2012) 17 *Journal of Conflict and Security Law* 212, 212-13.
[82] See James A. Green, 'Disasters Caused in Cyberspace' in Susan C Breau and Katja L H Samuel (eds), *Research Handbook on Disasters and International Law* (Edward Elgar, forthcoming 2016); and Michael N Schmitt, 'In Defense of Due Diligence in Cyberspace' (2015) 125 *Yale Law Journal Forum* 68.
[83] See Mary Ellen O'Connell, 'Cyber Security without Cyber War' (2012) 17 *Journal of Conflict and Security Law* 187, 204-205.
[84] *Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v United States of America)* (merits) [1986] ICJ Rep 14.

within the remit of Article 2(4).[85] To revisit those in considerable detail would undoubtedly detract from the more important discussion regarding autonomous responses. Therefore, the authors propose to confine themselves to highlighting Matthew Sklerov's work regarding cyber attack assessment rather than delving further into already well documented analysis.[86] It should be added that Sklerov himself draws heavily on Michael Schmitt's seminal works on cyber attacks.[87]

Broadly speaking, Sklerov sets out three categories / approaches, while contemplating any defensive strategy in response to a cyber attack.[88] First, is what he deems as an 'instrument-based approach'—this questions whether the damage caused (in this case by cyber operations) could only have been previously achieved by kinetic means.[89] Secondly, is what Sklerov refers to as an effects or consequence based approach—as its name suggests the means or mode of attack is irrelevant and of greater importance is the effect suffered by the victim state.[90] Finally, is a strict liability approach whereby all cyber attacks against CNIs are treated as armed attacks because of the severe nature of the consequences that ensue.[91]

It is highly tempting to take the view that the second and third approaches are fairly similar in substance. However, and as Sklerov himself concludes, one can therefore take the view that no matter which model is used, cyber attacks can constitute an armed attack.[92] This is perhaps somewhat simplistic and as Green would quickly point out, one must recognise whether a cyber-attack does (or can) constitute an instance of 'force', 'intervention' and/or an 'armed attack' – differing concepts requiring different thresholds.[93] For the purposes of this article, perhaps the conclusion needed to be reached sooner rather than later is simply that cyber-attacks with *non*-kinetic results may qualify as a use of force and trigger the armed attack threshold thereby allowing a state to lawfully respond in self-defence.[94]

---

[85] In particular, see for example, Green (n 6) at 98-107. In addition, to Matthew J Sklerov, 'Solving the Dilemma of State Responses to Cyberattacks: A Justification for the Use of Active Defenses against States Who Neglect Their Duty to Prevent' (2009) 201 *Military Law Review* 1, particularly at 54-5, Michael N Schmitt 'Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework' (1999) 37 *Columbia Journal of Transnational Law* 885. See also Roscini (n 6), Waxman (n 6), Hathaway, Crootof, Levitz, Nix, Nowlan, Perdue and Spiegel (n 6), and Dinniss (n 14).

[86] Sklerov (n 85) 54-55.

[87] Schmitt (n 85).

[88] Sklerov (n 85) 54-55.

[89] Ibid.

[90] Ibid.

[91] Ibid.

[92] Ibid.

[93] James A Green, *The International Court of Justice and Self-Defence in International Law* (Hart, 2009), 31-3.

[94] See, for example, Jack Goldsmith, 'How Cyber Changes the Laws of War' (2013) 24 *European Journal of International Law* 129, 133; Vida M Antolin-Jenkins, 'Defining the Parameters of Cyberwar Operations: Looking for Law in All the Wrong Places?' (2005) 51 *Naval Law Review* 132, 155; and Stephanie G Handler, 'New Cyber Face of Battle: Developing a Legal Approach to Accommodate Emerging Trends in Warfare' (2012) 48 *Stanford Journal of International Law* 209, 229. See also Dinnis (n 14) 81, who takes the view, while referring to cyber attacks on Estonia and Georgia, that the attacks do not go beyond the equivalent of a frontier incident.

C. **Self-Defence Applicable to Cyber Operations**

While the prohibition contained in Article 2(4) of the United Nations Charter remains absolute, there exist two, well known 'exceptions' to this default position: self-defence and collective security. Since the focus of this article is entirely on the former in terms of automated self-defence, the present discussion will not entertain any foray into collective security. Given the considerable wealth of commentary already dedicated to both visiting and revisiting a state's inherent right to invoke self-defence under customary international law, such discussions will be similarly avoided.

The law governing a state's inherent right of self-defence fuses pre-existing Charter Law in the form of International Custom and Article 51 of the UN Charter.[95] Article 51 is explicitly in is requirement that state must have suffered an 'armed attack' (or at the very least be faced with a sufficiently serious and imminent threat of suffering an armed attack).[96] Neverthelss, the text of Article 51 provides no guidance to the term of art, 'armed attack' and its applicable threshold. Instead, further explanation is distilled from the ICJ's judgment in the *Nicaragua* case[97] and from commentators alike, to mean that an armed attack should be defined as: 'the most grave form of the use of force'—a qualitatively grave use of force—beyond a use of force simpliciter.[98]

Once a state has suffered an armed attack, the lawfulness of its response is regulated by the two key parameters: necessity and proportionality. The origins of necessity and proportionality were espoused in the seminal correspondence between the then US Secretary of State Daniel Webster, and his British counterpart Lord Ashburton with regards to and forming part of the *Caroline* incident.[99] Daniel Webster's formulation required that in order for a state to lawfully invoke self-defence it would need to:

> '[S]how a necessity of self-defence, instant, overwhelming, leaving no choice of means, and no moment for deliberation. It will be for it to show, also, that . . . [it] did nothing unreasonable or excessive; since the act, justified by the necessity of self-defence, must be limited by that necessity, and kept clearly within it.'

[95] James A. Green and Francis Grimal, 'The Threat of Force as an Action in Self-Defense Under International law' (2011) 44 *Vanderbilt Journal of Transnational Law* 285-329, at 299.
[96] Don W. Greig, 'Self Defence and the Security Council: What Does Article 51 Require? (1991) 40(02) *International and Comparative Law Quarterly* 366-402. NB: Article 51 of the UN Charter remains silent as to imminence.
[97] *Nicaragua* (n 84).
[98] Green & Grimal (n 95) p 300, Avra Constantinou, *The Right of Self-Defence Under Customary International Law and Article 51 of the United Nations Charter* (Doctoral Dissertation, University of Nottingham, 1996).
[99] Letter from Daniel Webster to Henry S. Fox (Apr. 24, 1841), in 29 *British and Foreign State Papers* (1841-42), 1129-39 (1857).

The principles of necessity and proportionality; both of which are intertwined, are distilled from the Webster formulation.[100] Deriving from the Webster formulation / *Caroline* formula, the modern interpretation of necessity dictates and questions whether it was reasonable to use force as a last resort and, if non-forcible measures were a reasonable alternative in the circumstances, that they were explored/exhausted—a measure therefore of last resort.[101] The proportionality meanwhile, requirement dictates that the "force employed must not be excessive with regard to the goal of abating or repelling the attack".[102] Green and Grimal both remark that a state's response need not actually mirror the initial attack, numerically speaking. If state A fires 10 missiles at state B, then state B is not obliged under the concept of proportionality to respond with a volley of 10 identical missiles.[103]

At this juncture, it is important to distinguish the lawfulness of a defending state's action taken during an on-going armed attack–the so called 'cumulative effect', as phrased by Garwood-Gowers,[104] and instances where force is used once the armed attack has ceased. In the context of the former (an on-going armed attack), according to Green, the position is that the responding State is placed under a temporal restriction—there must be a reasonable temporal proximity between the victim State's response and the armed attack itself.[105] Undeniably, and Green himself is the first to concede and highlight that the 'reasonableness' parameter is somewhat nebulous and imprecise.[106] Accordingly, green suggests this area is open to interpretation along the lines of 'a context-specific appraisal of the various factors that may delay a self-defence action: intelligence gathering, initial resort to negotiation, geographical disparity, and so on'.[107] The authors of this present article are of the view that an 'overly tardy' response may negate the necessity requirement—the longer a state waits before responding, the more difficult it is to reconcile with the 'last

[100] Green & Grimal (n 95). See generally James A. Green 'Docking the Caroline: Understanding the Relevance of the Formula in Contemporary Customary International Law Concerning Self-Defence' (2006) 14 *Cardozo Journal of International & Comparative Law* 429.

[101] D Kretzmer, 'The Inherent Right to Self-Defence and Proportionality in *Jus Ad Bellum,*' *European Journal of International Law* Vol 24 No. 1 (2013) 235-282. J Gardham, *Necessity, Proportionality, and the Use of Force by States* (Cambridge University Press 2004); S Etezazian, 'The Nature of the Self-Defence Proportionality Requirement,' Journal on the Use of Force and International Law (2016) 1-30; See also T Christodoulidou and K Chainoglou, 'The Prnciple of Proportionality From a Jus Ad Bellum Perspective,' in Marc Weller (ed.) *The Oxford Handbook of The Use of Force in International Law* (Oxford Publishing 2015). Green & Grimal (n 95The).

[102] See Constantinou (n 98) 159-61; Gamal Moursi Badr, 'The Exculpatory Effect of Self-Defense in State Responsibility' (1980) 10 *Georgia Journal of International & Comparative Law* 1; David Kretzmer, 'Killing of Suspected Terrorists: Extra Judicial Executions or Legitimate Means of Defence?' (2005) 16(2) *European Journal of International Law* 171-212.

[103] Green & Grimal (n 95) 301. See Judge Higgins's Dissenting Opinion, in Advisory Opinion *Legality of the Threat or Use of Nuclear Weapons* [1996] ICJ Rep 226 at para 5. See also generally, David Kretzmer, 'The Inherent Right to Self-Defence and Proportionality in Jus Ad Bellum' (2013) 24(1) *European Journal of International Law* 235-282.

[104] *See generally* Andrew Garwood-Gowers, 'Self-Defence Against Terrorism in the Post-9/11 World' (2004) 4 *Queensland University of Technology Law and Justice Journal* 1.

[105] See James A. Green 'The *Ratione Temporis* Elements of Self-Defence' (2015) 2(1) *Journal on the Use of Force and International Law* 97-118.

[106] Ibid.

[107] Ibid.

resort' criteria required for the necessity element. Therefore, an overly tardy response may render the action unlawful.

Undeniably, there is also a strong case to be made that a state's response to a cyber attack (which, may amount to an actual 'armed attack') could be via measures falling short and not giving rise to an actual use of force.[108] This also leaves open the more problematic question as to what typology of reaction is envisaged against a cyberattack (providing it meets the armed attack threshold)? Different typologies of response could of course encompass a more traditional 'kinetic reaction', or a reaction in cyberspace falling short of a grave use of force.[109] Both of which, may be just as effective in the strategic sense of abating the initial attack / strike.[110] In the case of the latter, one would also need to consider the availability of 'countermeasures' or other applicable defences within the context of precluding wrongfulness such as necessity, distress or force-majeure—the advantage of which, is that the perennial problem of attribution is thus negated.[111] However, the present authors are of the view that while such responses falling short of actual force may well be desirable in terms of justification (of a lawful response), the non-forceful element may not be sufficient to abate or repel a further attack.

From a proportionality perspective, clearly certain questions require further investigation. First, if a state is invoking its inherent right of self-defence, would the response need to be limited to that of a cyber nature or could it also be kinetic? Under a strict interpretation of proportionality, one could argue that the means of response actually alters very little. The consideration would still be whether the response be it cyber or kinetic does not exceed the defensive necessity of abating or repelling a further attack. Certain scenarios may invite one or other responses from a strategic perspective but in terms of application, there is nothing inherently wrong with leaving the door ajar for both. More controversial however, is to whom the response is directed against rather than the mode. Would for example the response need to be targeted against a state's cyber capabilities rather than its other military assets? Or, would a cyber attack by zombie computers 'invite' a lawful and proportionate response of destroying innocent civilian computers so as to thwart the attack? Undeniably, the latter runs the risk of exceeding the careful framework required in order to maintain a proportionate response of abating or repelling a further attack—another reason why automated responses have inherent deficiencies.

D. The Case for Re-Calibration of Temporal Limits[112]

[108] The authors are grateful to Professor Dr Tom Ruys for his helpful contribution and suggestion on this point.
[109] Ibid. Since this issue has been dealt with elsewhere within the literature by Dinniss (n 14), this article confines itself to exploring its original remit and will as such, not delve further into the typology.
[110] Ibid. Ruys (n108).
[111] Ibid.
[112] N.B. Some of views expressed in this section follow a similar trajectory to those expressed by Grimal in his article: 'Missile Defence Shields: Automated and Anticipatory Self-Defence?' (2014) 19(2) *Journal of Conflict & Security Law* 317-339.

While the actual cyber attack may take place within nano-seconds, any lawful self-defence action can ensue once attribution has been established. In other words, the temporal space for a self-defence action against a cyber attack should be extended to allow for attribution to be completed to the satisfaction of the defending state. Self-defence against a traditional strike platform places the responding state under a temporal restriction requiring a reasonable temporal proximity between the victim State's response and the armed attack itself. Unlike a response against a conventional attack, a response in self-defence to a cyber attack after a lapse of twelve to eighteen months cannot categorically be ruled out due to difficulties in detection and attribution of the attack to a particular state.

However, the necessity element is therefore also being stretched—the greater the time lapse between attack and response, the more it affects the state's ability to fulfil the necessity requirement of self-defence. For example, it is more difficult for a state to argue that it is acting out of last resort, that it has exhausted all non-forceful measures, and that it will be wholly unreasonable to expect a non-forceful response. It can be argued that an immediate response could be ruled out and any self-defence response can only take place after a clear determination of the attribution of act and actor.[113] This may take months to achieve, or may not even be possible to achieve for a long period of time thereby making any self-defence action purposeless. Understandably, whilst this conceptual recalibration may be desirable, the practical consequences may not be. This leaves open the possibility for abuse by taking a legitimate action in self-defence and turning it into an unlawful reprisal.

In a traditional armed/kinetic attack, there is certainly the notion that to a limited degree, states can extend 'their response in self-defence beyond the moment where the attack being responded to terminated'.[114] For the reasons stated above the response time available for a state which has come under a cyber attack will have to be stretched, or recalibrated to give space to the affected state. In certain instances, a state may not even be aware it has come under a cyber attack.[115] Green and Garwood-Gowers postulate (independently) that when a state comes under a kinetic attack there exists a 'dual' or 'cumulative effect' argument whereby that state not only needs to respond to the previous attack but be guarded against a future attack.[116] This 'dual' or 'cumulative effect' alluded to by Green and Garwood-Gowers may have to be tweaked to first look at beefing up the defences from any future attacks while one investigates the source, origin, and key players to the attack for purposes of attribution, as any reprisals without clear attribution of act and actor is bound to backfire and be counter productive.[117] The investigation could also take a lengthy period requiring the recalibration of the temporal limits to complete any investigation before devising a suitable response.

---

[113] See chapter IV for a detailed discussion on attribution.

[114] Ibid. And, as helpfully signposted by the anonymous reviewer, this would cover 'Crimea-type' scenarios whereby the 'defending state' has since been occupied.

[115] See for instance, the incident involving Stuxnet, where Iran was not aware that it had come under a cyber attack, and continued to address the problem differently.

[116] Ibid. See also Garwood-Gowers (n 105).

[117] Motives and attribution are discussed in part IV of this article.

While discussing the possibility of establishing a framework for automatic self-defence in cyberspace, it is imperative to look at some of the arguments arising under anticipatory self-defence under traditional kinetic warfare. Since the ICJ's refusal to reject the possibility of anticipatory self-defence in the *Nicaragua Case*, anticipatory self-defence has remained highly controversial amongst academics and states alike.[118] Here, the controversy hinges on the lawfulness of a forcible response against an imminent threat of force rather than an actual use of force.[119]

Taking action against a *mere* threat of force is doubly problematic: previous cyber attacks have invariably been perpetrated *via* non-state actors (with state backing—hard to prove), hacktivists and others. The examples referred to in part II clearly demonstrate that states were unable to gauge an 'imminent threat of cyber attack/imminent cyber threat' beforehand to be able to respond by force of any kind. In the absence of a clear cut attribution, it may be an extremely difficult task to respond to a threatened cyber attack. At this juncture, one can state that the same arguments available to a state which faces an imminent threat of force (kinetic) in order to exercise a forcible response may not be available to a state which fears an imminent cyber attack. This is because it is an extremely difficult task to pin point with accuracy the origin of a cyber attack /perpetrator of the attack with absolute certainty with the current level of technical expertise at our disposal.

If we are to take the position that self-defence against any attack, whether kinetic or cyber, will have to strictly meet the parameters of international law/Charter Law, it will then require a discussion of armed attack as understood under the provisions of Article 51. Would a state need to have suffered an 'armed attack' as understood under the language of Article 51, or could it rely upon the customary position set out by the *Caroline* formula—enabling a state to lawfully use anticipatory force against an imminent cyber threat?[120] Debate also surrounds the terminology used by scholars.[121] The position taken by this author

---

[118] Nicaragua (n 84). See Green & Grimal (n 95) 287. See also Jackson N. Maogoto, *Battling Terrorism: Legal Perspectives On The Use Of Force And The War On Terror* (Ashgate Publishers, 1st Edition, 2005) 111–149, where Maogoto gives a useful overview of the main arguments concerning this issue and provides a survey of the vast literature. See also Christine Gray, 'The US National Security Strategy and the New "Bush Doctrine" on Pre-emptive Self-Defence' (2002) 1 *Chinese Journal of International Law* 437, 438 (describing the 'radical new doctrine of international law on the use of force'); Christopher Greenwood, 'International Law and the Pre-Emptive Use of Force: Afghanistan, Al-Qaida, and Iraq' (2003) 4 *San Diego International Law Journal* 7, 8 (noting that some commentators have called for amendment to the UN Charter); Christian M. Henderson, 'The 2006 National Security Strategy of the United States: The Pre-Emptive Use of Force and the Persistent Advocate' (2007) 15 *Tulsa Journal of Competition & International Law* 1, 2 (characterizing the 2006 reassertion of the doctrine of pre-emptive military action as "surprising"); Abraham D. Sofaer, 'On the Necessity of Pre-Emption' (2003) 14 *European Journal of International Law* 209, 210 (noting that traditional deterrence is ineffective against terrorists); see generally Miriam Sapiro, 'Iraq: The Shifting Sands of Pre-Emptive Self-Defence' (2003) 97(3) *The American Journal of International Law* 599 (arguing that the United States should refine its position on the preemptive use of force).

[119] Ibid.

[120] Green (n 100) 463-73.

[121] Christine D Gray, *International Law and the Use of Force* (Oxford University Press 3rd ed. 2008) 211-212.

both here and elsewhere is that anticipatory self-defence refers to action taken in response to an imminent threat; pre-emptive self-defence, meanwhile, is action taken against a latent and temporally remote threat.[122] The major hurdle in trying to adapt the customary position of international law, *i.e.*, *Caroline* formula to a situation where a state fears an imminent threat of a cyber attack is not knowing who the attacker is and what is being sought to be targeted. In simple terms anticipatory self-defence follows the wording of the *Caroline* formula—a state must respond to a threat which leaves "no moment for deliberation".[123]

This principle may not sit comfortably to support an action in anticipatory self-defence where a cyber attack is seen as imminent, as it may be extremely difficult to demonstrate that the cyber threat leaves very little time to deliberate and one needs to strike in self-defence. Here, a perceived imminent cyber attack appears to be a remote argument as in a traditional kinetic attack a state would not only know its enemies and allies, but also their capabilities and possible movements through surveillance, etc., before deciding to act.

In *Nicaragua,* the ICJ adopted the following position:

> [I]n the circumstances of the dispute now before the Court, what is in issue is the purported exercise by the United States of a right of collective self-defence in response to an armed attack on another State. The possible lawfulness of a response to the imminent threat of an armed attack which has not yet taken place has not been raised.

According to Gill the logical interpretation of the ICJ's pronouncement is that for anticipatory action to be lawful it would have to be taken against a threatened armed attack.[124] But, state practice meanwhile appears to be predicated purely on the concept of imminence.[125] In essence, the threat posed

---

[122] Green & Grimal (n 95). See Constantine Antonopoulos, 'Force by Armed Groups as Armed Attack and the Broadening of Self-Defence', (2008) 55(02) *Netherlands International Law Review* 159, 172; and Niaz A. Shah, 'Self-Defence, Anticipatory Self-Defence and Pre-Emption: International Law's Response to Terrorism' (2007) 12(1) *Journal of Conflict & Security Law* 95, 111.

[123] *Nicaragua* (n 84).

[124] Terry D. Gill, 'The Law of Armed Attack in the Context of the Nicaragua Case' (1988) 1 *Hague Year Book of International Law* 30, 35.

[125] Green & Grimal (n 95) 105. The best example of this followed the 1981 Israeli attack upon the Iraqi Osiraq nuclear reactor, after which Israel explicitly justified its action as anticipatory self-defence. See UN SCOR 36th Sess., 2288th mtg. at 79-84, UN Doc. S/PV.2288 (19 June 1981) ("Israel had full legal justification to exercise its inherent right of self-defence . . .."); Gray (n 128) at 115. In doing so, Israel itself argued that the danger posed by the Iraqi reactor was imminent. See UN SCOR 36th Sess., 2288th mtg. at 102, UN Doc. S/PV.2288 (June 12, 1981) ("We [Israel] waited until the eleventh hour after the diplomatic clock had run out . . .."). States almost universally condemned the action, but, notably, most states did so on the basis that the threat to Israel was, contrary to what Israel had claimed, not imminent. See, e.g., UN SCOR 36th Sess., 2288th mtg. at 28-30, UN Doc. S/PV.2288, (June 19, 1981) (noting that while Israel may have legitimately felt threatened, there were still non-military solutions available); UN SCOR 36th Sess., 2288th mtg. at 44-47, UN Doc. S/PV.2288 (June 16, 1981) ("Today the Israelis attack Baghdad for having a nuclear reactor centre that was described by the . . . IAEA . . . as 'peaceful nuclear facilities.'"); UN SCOR 36th Sess., 2288th mtg. at 53-56, UN Doc. S/PV.2288 (June 15, 1981) (referring to the air raid on Iraq's capital as an "unprovoked" act of terrorism). Of course, a

will need to be qualitatively grave (a threatened armed attack) and also imminent in order for self-defence to be lawfully invoked,[126] by the state pleading the case of imminence. Pre-emptive self-defence on the other hand, stretches the 'elasticity' of imminence to breaking point,[127] which is well captured in the 'Bush Doctrine' in the United State National Security Strategy 2002, where the US effectively removed the imminence requirement, *i.e.*, action will be taken against a latent threat that may or may not materialise at some indeterminate point in the future.[128] Under Bush Doctrine, any pre-emptive self-defence action taken will be seen as being perfectly lawful—a proposition rejected by states and scholars alike.[129]

Dinstein's discussion of a hypothetical attack by American Forces against the Japanese fleet, so as, to prevent the attack on Pearl Harbour in December 1941, and his concept of interceptive self-defence are of particular relevance to our discussions,[130] as it presents a scenario where the temporal element of attack,

---

number of other states argued that the action was unlawful because self-defence against a threat is unlawful per se; for example, the Soviet Union referred to such actions as "the law of the jungle."

[126] See Shah (n 122) 101–04, 111–19 (describing the gravity and immediacy of the threat required to justify self-defence under international law).

[127] The authors are grateful to Robert Barnidge Jr. for the following observation. John Brennan during his tenure as Obama's homeland security advisor argued that practice also supports a more flexible understanding of imminence.

[128] The United States stated that it would resort to the pre-emptive use of force "even if uncertainty remains as to the time and place of the enemy's attack." See *The National Security Strategy of the United States of America* (2002), available at <http://www.state.gov/documents/organization/63562.pdf> (accessed 12 December 2016). This position was restated, essentially unmodified in 2005 and 2006. See *The National Security Strategy of the United States of America* (2006) 18, 23, available at <http://www.presidentialrhetoric.com/speeches/nss2006.pdf> (accessed 12 December 2016z; US Department of Defence, The National Defence Strategy of the United States of America (2005) 9-12, available at <http://archive.defense.gov/news/Mar2005/d20050318nds1.pdf> (accessed 12 December 2016).

[129] Green & Grimal (n 95). See, for example, the categorical rejection of the notion of pre-emptive attack by the Non-Aligned Movement in the declaration that emerged from the Fourteenth Summit of Heads of State or Government of the Non-Aligned Movement. Non-Aligned Movement, Final Report Covering the 14th Conference of Heads of States or Governments of the Non-Aligned Movement (11-16 September 2006) para 22.5, available at <http://cns.miis.edu/nam/documents/Official_Document/14NAMSummit-Havana-Compiled.pdf> (accessed 12 December 2016). See for example, Tarcisio Gazzini, *The Changing Rules on the Use of Force in International Law*, (Juris Publishing, 2005) 174, 238. In the authors opinion a state's practice '…is neither quantitatively nor qualitatively consistent enough to affirm the existence of a right to anticipatory self-defence, a development that would stretch beyond recognition the notion of self-defence itself.' See also Greenwood (n 118) 12–16, where the author notes that the right to the exercise of '…anticipatory self-defence is confined to instances where the armed attack is imminent'; and Sapiro (n 118) 599–603. The author opines that although it is possible to interpret the law to permit defensive action in the face of imminent threat, it will be not only difficult but also dangerous to stretch it further.

[130] See Yoram Dinstein, *War Agression and Self-Defence* (Cambridge University Press, 5th Edition, 2011) 203-204. A similar discussion / example has also been used by Cassese regarding Anticipatory action. As Cassese writes, the rationale is a strong meta-legal argument to prevent in McDougall's words a state becoming a 'sitting duck' to impending military attacks. Cassese provides the hypothetical scenario of the US Pacific Fleet sinking the Japanese carrier *en route* to Pearl Harbour in 1941 as an example. See Antonio Cassese, *International Law* (Oxford University Press, Oxford 2nd Edition 2005) 308.

preparedness and interceptive self-defence/response is carefully deconstructed. Within the modality of self-defence, Dinstein's terminology of 'interceptive self-defence' fully encapsulates the essence of the role of a CNI intercepting an on-going attack—the 'countering of an armed attack which is already in progress'.[131] When Dinstein's notion of 'interceptive self-defence' is brought into the cyber realm to thwart any imminent cyber attack one encounters difficulties, most significant of which is with regard to the different temporal limits present in the two dimensions of warfare played out *viz*., an air attack and a cyber attack.

As regards an air strike is concerned, the *mere* target acquisition and 'locking on' by a fighter jet could constitute an armed attack (albeit in progress), and according to Dinstein, a 'timely response' against the fighter jet would constitute interceptive self-defence.[132] Within the 'fifth dimension', one could analogise that the CNI has detected some abnormality in the data traffic and the CNI consequently automatically 'intercepts' a potential cyber attack. Unlike a missile-defence shield that intercepts a missile that is already programmed with a payload and whose launch is undeniably imminent, the point of interception is much less clear cut.

Whereas with the cyber interception, such activity or interception is taking place within considerably shorter temporal limits (nano seconds) and also in cyber-space—a battlefield which, is much less tangible. When data packets are fired in the battlefield terrain of computer networks, aiming to breach secure internet protocols, they assume a different time frame to that of a conventional battlefield, where bullets and missiles are fired in real space against real targets.

It is well known that botnets (zombie computers) are extensively used in cyber attacks, and the uneven spatial distribution of infected computers across the internet, specifically on a limited number of 'unclean' or 'zombie-friendly' networks pose a major problem.[133] Botnets, which could be 'turned into digital weapons,'[134] are normally located in different jurisdictions and connected to different servers and internet providers. Most worryingly, the botnets studied over a period of time demonstrated an impressive attack capability, and further, like real-world armies, they were capable of coordinating their efforts with other botnets.[135]

One should note that much of Dinstein's discussion of Pearl Harbour relies implicitly on an imminent threat in terms of affecting the lawfulness of his

---

[131] Ibid, 204.

[132] Ibid.

[133] The authors are again grateful to Professor Dr Tom Ruys for his suggested inclusion of the downing of Iran Air Flight 655 as an example of the inherent dangers of 'trigger happy' interceptive action.

[134] Oliver Thonnard, Wim Mees and Mark Darier, 'Behavioral Analysis of Zombie Armies' in Christian Czosseck and Kenneth Geers (eds), *The Virtual Battlefield: Perspectives on Cyberwarfare* (IOS Press, 2009) 191-210.

[135] Ibid. The finding from the study carried out by the authors are interesting, such as a) botnets/zombies demonstrate extraordinary resilience on the internet, with survival times going up to several months; b) there is high degree of coordination among zombies; and c) the presence of a large proportion of home users' machines with high-speed Internet connections among the bot population. These findings obviously present a worrying picture, as at a given point in time there could millions of computers spread across a region that are infected and compromised, and ready to be part of a cyber attack.

hypothetical scenario.[136] He then envisages 3 types of hypothetical scenarios,[137] which are as follows:

1) The shooting down of a Japanese Type 99 Carrier Bomber just prior to it attacking Pearl Harbour. The Bomber would have left the carrier and would be inbound and poised to drop its ordnance. According to Dinstein, such an attack would be lawful—once the aircraft have been launched from the carrier, there can be no doubt that an armed attack is underway and that the other side has "committed itself to an armed attack in an ostensibly irrevocable way".[138]
2) The sinking of the Japanese Fleet prior to the launch of any aircraft poised to attack on the US's Pacific Naval Base and Pearl Harbour. This is much more problematic and as Dinstein concedes, lawfulness would hinge on real time concrete data visibly demonstrating that Pearl Harbour would be subjected to an imminent attack; reminiscent perhaps of satisfying the '*Caroline* criteria'.
3) An attack by the US against the Japanese fleet prior to it setting sail or during war gaming manoeuvres. This is very much along the pre-emptive lines and as Dinstein rightly concludes, would be undeniably unlawful.

Therefore, under the Dinstein sense or 'necessity', a reasonable interpretation of interception would be along the *Caroline* incident lines.[139]

When one takes a closer look at the three possible actions presented by Dinstein to thwart an impending air attack (modelled on the Pearl Harbour strike), it becomes clear that firstly, there is unequivocal information about the attacker, the mode of attack and the weapons used, and sufficient time to intercept/press into service one of the three possible actions presented. Concretely, in cyber attacks there is much less room for manoeuvre in terms of timings because the detection of such an attack is much more problematic. Unlike missiles which, are in the 'free flight phase' or ideally at the 'boost phase' (noting that it is difficult to determine exact trajectory in this phase), it is much more difficult to ascertain both the point of origin and the source of the attack in the cyber attack in the cyber realm. For both missile interception and a cyber response, interception would probably fall within the realm of necessity.

However, for the necessity threshold to be triggered in the cyber realm one would have to conclude (in a very compressed timeframe) that the system is indeed under attack and that there are no alternatives available in order to defend the system. In other words, a state is acting anticipatorily—something that the Court in *Nicaragua* did not dismiss outright in paragraph 35 and, of course, if one accepts a more general right of anticipatory self-defence under international law.[140] A response under those set of circumstances against a considerable surge in data would arguably fall within the necessity

[136] Dinstein (n 130) 204.
[137] Ibid.
[138] Ibid.
[139] See Green (n 100).
[140] See for example, Constantine Antonopolos, 'Force by Armed Groups as Armed Attack and the Broadening of Self-Defence' (2008) 55 *Netherlands International Law Review* 159-180.

requirement. A cyber attack is carried out using data streams (which the internet is based on), and any unusual data surge need not necessarily indicate/translate into an impending cyber attack. Importantly, the data stream is used to carry all forms of computer network attacks, which vary widely,[141] and could also be manipulated to mask or hide the origin of the stream and hence the attacker to thwart any detection. This aspect will be further discussed in part IV under motives and attribution.

IV. AUTONOMOUS SELF-DEFENCE IN CYBER OPERATIONS

As noted in the introduction, the authors wish to introduce and explore the novel concept of automated self-defence in the cyber realm as a means of explicating an action which, is 'machine guided', that is, devoid of human involvement and hence 'automatic'. The programming of any machine (to this day at least) is undertaken by human beings. However, the cause for concern with regards to automated self-defence is that the 'machine is calling the shots' (sic)—the lack of human control is concerning for the application of these criteria. NATO has defined Computer Network Defence as *'Actions taken through the use of computer networks to protect, monitor, analyze, detect and respond to unauthorized activity within information systems and computer networks'*.[142] The above definition can be used as a working model to develop an argument for an automated Cyber Defence Shield (CDS). The definition limits/presupposes that any defensive activities to be through the use of computer networks to counter any unauthorized activity within the cyber realm.

The purpose of this section is to study in greater detail the threshold of response, that is to say, against which types of actions or threats are automated responses calibrated to? The format for analysis in this section is as follows: Part

---

[141] This would include gaining access to a computer system (to acquire control over it), transmitting viruses to destroy or alter data, using logic bombs that sit idle in a system (to be triggered off later), inserting worms that reproduce themselves upon entry into a system resulting in overloading the network, and employing sniffers to monitor and/or seize data. See Michael N Schmitt, "Wired Warfare: Computer Network Attack and Jus ad Bello' (2002) 84(846) *Revue Internationale de la Croix-Rouge/International Review of the Red Cross* 365-399, 367. See also Dinniss (n 14) where the author highlights IP spoofing and backdoor payloads. High value data encryption is used to secure data and to build a sound defence against data theft. Public key cryptography, or asymmetric cryptography is used in encrypting the Transmission Control Protocol / Internet Protocol (TCP/IP) communication between network end points. The objective of encryption is to make it impossible to protect the confidentiality of digital data stored on computer. Encrypted data also provides confidentiality while also ensuring authenticity of the provider. Public key cryptography (or asymmetric key cryptography) which uses a number of algorithms for the purposes of securing data is the most popular one.

[142] NATO publication 3000 TI-3/TT-1162. See also Luc Beaudoin, Nathalie Japkowicz and Stan Matwin, 'Autonomic Computer Network Defence Using Risk State and Reinforcement Learning' in in Christian Czosseck and Kenneth Geers (eds), *The Virtual Battlefield: Perspectives on Cyberwarfare* (IOS Press, 2009 238-248. The authors talk about Computer Network Defence (CND) is concerned with the active protection of information technology infrastructure against malicious and accidental incidents. They also opine that CND requires *an automated controller with a policy, which selects the most appropriate action in any undesired network state*. Due to the *complexity and constant evolution of the CND environment, a-priori design for an automated controller is not effective*.

A will provide an overview in terms of threats of force. This is central to the argument because to unravel the nature of the automated response it is necessary to understand what the CDS is responding to in terms of its calibration.

Is it a threat of force or an *actual* use of force? Can there be a threat of force in cyber space, as in the case of an air strike or missile strike? Is the system at large to be viewed as the target of such a cyber threat, as opposed to a specific target in conventional warfare? The second of the questions is raised here due to the peculiar landscape of cyber space where internet serves as the medium through which not only information is exchanged through data streams but also most of the activities of the state (both military and civil) are conducted,[143] making it a most desired target for an attack.

The instances of cyber attacks studied under part II encapsulates the different ways in which an attack could be carried out on both civil and military targets. Part B will explore whether, and to what extent, the response by a CDS fits within the threshold of necessity and proportionality.

A. Threats of Force

Before proceeding to define and consider threats of force, it is helpful to draw on Dinstein's Pearl Harbour scenario by way of explanation – especially to scenarios 2 and 3. To recall, scenario 2 discussed the sinking of the Japanese Fleet prior to the launch of any aircraft poised to attack the US's Pacific Naval Base and Pearl Harbour. Scenario 3 envisaged an attack by the US against the Japanese fleet prior to it setting sail or during war gaming manoeuvres. Taking a slightly more controversial line, one can argue that 'interception' (to use Dinstein's terminology), in both scenarios could also be predicated on a response to a threatened cyber attack[144] and not an actual cyber attack, and thus in CDS terms may affect the way in which any interception operates.

Conceivably, 'interception' in scenario 2 could be against an imminent grave threat of cyber attack rather than an 'actual cyber attack' (something Dinstein appears to implicitly allude to—the authors of this article transpose into the cyber realm). Here, there is no reference as in scenario 1 for the US 'to regard the Japanese armed attack as having commenced'.[145] As mentioned earlier, one should still take into consideration the temporal recalibration that is required to apply these principles to the cyber realm, as the response time will be severely limited when one talks about interception in the cyber realm. Also once the attack has commenced there is little or no time to intercept, as things happen in a nano-second in the cyber realm.

---

[143] The internet has moved from being a platform to access and exchange information while still it was in its infancy in the latter part of the twentieth century, to a realm to conduct day-to-day affairs (civil and military) in the twenty first century. In current day terms, the more a society could do on the internet, the more modern it is as a state. Simply put, a modern state is fully 'wired'.

[144] Where patterns of data exchanged, peculiar data surges, etc. originating from a certain region strongly indicates that it is highly likely that state A was intending to carry out a cyber attack, and state B is constrained to consider its response options, including the use of the CDS.

[145] Dinstein (n 130) 203.

Scenario 3 appears in threat terminology at least to be against a non-imminent and latent. It is important not to overplay this discussion *vis-à-vis* threats, imminent threats, etc., but in order to fully understand the lawfulness of actions taken by a CDS one must be mindful as to how threats operate in order to appreciate that interception against a non-imminent latent threat, as opposed to an actual cyber attack may yield very different results in terms of both necessity and proportionality.

It is important to note that threats of force remain a nebulous concept under international law. Although they are prohibited, they still remain undefined by Article 2(4) of the United Nations Charter,[146] which runs as follows:

> 'All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations.'

The prohibition against the threat of force has also been restated in the form of soft law declarations in *1970 Declaration on the Principles of International Law Concerning Friendly relations and Cooperation Among States* and the *1987 Declaration on the Enhancement of the Effectiveness of the Principle of Refraining from the Threat or Use of Force in International Relations.*[147]

[146] UN Charter Article 2, para 4. See Green & Grimal (n 95); See also Dino Kritsiotis, 'Close Encounters of a Sovereign Kind' (2009) 20(2) *European Journal of International Law* 299-330. It is generally accepted that the prohibition of the use of force is also universally binding under customary international law. See for example, Michael Bothe, 'Terrorism and the Legality of Pre-Emptive Force' (2003) 14(2) *European Journal of International Law* 227-240, 228 where the author observes that the '…prohibition of the use of force is a valid norm of customary international law...'; Hermann Mosler, 'The International Society as a Legal Community' (BRILL, 1974) 140 *Collected Courses of The Hague Academy of International Law* 283. Whether this is also true for the prohibition of the threat of force is debatable given the lack of clear articulation of the prohibition in state practice. It is also generally agreed in the literature that the prohibition of the use of force is a *jus cogens* norm (a peremptory norm of international law from which no derogation is possible). See for example, Alexander Orakhelashvili, *Peremptory Norms In International Law* (Oxford University Press, 2006) 50, where it is noted that the '…prohibition of the use of force by States undoubtedly forms part of *jus cogens*'. Some scholars have taken this further and argued that the prohibition of the threat of force is similarly a *jus cogens* norm. However, for the suggestion that the prohibition does exist in custom, see Nicholas Stürchler, *The Threat of Force in International Law* (Cambridge Press, 2007) 92–126. The observes that it is '…safe to conclude that article 2(4) of the UN Charter is *jus cogens* as a whole, without distinction to be made between the threat of force and the actual use of force'. However, it is our view that the peremptory status of the prohibition of the use of force is in fact debatable, and the prohibition of the threat of force is certainly not peremptory. See generally James A. Green, 'Questioning the Peremptory Status of the Prohibition of the Use of Force' (2010) 32 *Michigan Journal of International Law* 215 (regarding the peremptory status of the prohibition of the use of force); ibid at 225–29 (specifically regarding the peremptory status of the prohibition of the threat of force).

[147] Declaration on the Enhancement of the Effectiveness of the Principle of Refraining from the Threat or Use of Force in International Relations, G.A. Res. 42/22, U.N. Doc. A/42/22/766 (18 November 1987).

Commentators have posited that a threat may take a different guise—not necessarily something said but also something done albeit the archetypal threat remains a coded warning / ultimatum—comply or else.[148] For the purposes of the current investigation, the typology of threat is limited – if one were to accept that anticipatory self-defence is lawful, then the threat being responded to in self-defence must be a threatened armed attack, and moreover, the threatened armed attack must be imminent – such as a missile launch.[149] As Gill commenting on the judgment of ICJ in the *Nicaragua* case observes that there 'can be no doubt that an armed attack, or at any rate the threat of an armed attack, is an *absolute precondition* for the exercise of the right of self-defence'.[150]

The authors of this article maintain that a full assessment of a threat of force cannot be conducted without reference to strategic considerations. Strategic considerations help explain the practical distinction between an empty threat—made by a state that does not possess the means of carrying it out (which may well violate Article 2(4) but is 'tolerated') and a threat that is all too 'real'. Under traditional kinetic warfare, the threatening state is militarily capable of carrying out its threat, and the threat itself is both unlawful under Article 2(4) and intolerable in the eyes of the international community. The authors of this article maintain the view that recourse to Schelling's model set out in 'Arms and Influence' forms the basis of understanding the severity of a threat—particularly in terms of military appraisal, and helps clarify whether it is a grave threat of force.[151]

In order for the threat to be considered real/serious, the threatening state must possess the capability in terms of military platforms and strike force to deliver the payload. Also, that state needs to communicate its intention to its enemy that it will carry out the threat, and that threat must be credible. Within the context of CDS, the strategic considerations would considerably differ, as military capability and cyber capability do not equate—while the first of the two is more tangible and hence measurable, the second is not so tangible and hence not measurable. Clearly, a state that would have fired a missile, would have ticked all of the relevant boxes, but maybe not in the case of a cyber threat.

In terms of assessing the lawfulness of a threat of force (kinetic or otherwise), the present test under international law is the one put forward by the ICJ in the *Nuclear Weapons* advisory opinion.[152] Broadly-speaking, the test poses a retroactive test to the hypothetical question – *if* the threat of force were

---

[148] See Francis Grimal, *Threats of Force: International Law and Strategy* (Routledge, 2012). In particular refer to chapter 2, where it is posited that approaches range from categorisation and placing threats on a scale ranging from the innocuous to the extreme to examining the very purpose of the threat. For example, can non-verbal actions such as engaging in military exercises near another state's border fall within the remit of 2(4)? Or, is 2(4) solely concerned with verbal *ultimata* demanding compliance?

[149] Green & Grimal (n 95).

[150] Gill (124).

[151] Thomas C. Schelling, *Arms and Influence: With a New Preface and Afterword* (Yale University Press 2008).

[152] Internationaler Gerichtshof, 'Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion' (1996) *ICJ Reports* 226-593; Ian Brownlie, *International Law and the Use of Force by States* (Oxford: Clarendon Press, 1963). See the Introduction and Grimal (n 154) note Chapters 2 and 4.

carried out (in other words, if actual force and not threatened force were to be used), would it be lawful? If yes, then this would legitimise the prior threat. If not, if actual force would be deemed unlawful, then so would the threat that precedes it.

The question that needs to be posed for the current study is can the above template fit the requirement of response to an imminent cyber attack? A second question will be, can a state threaten another sovereign state with a cyber attack? The dangers of such an action can be multifold, as it could expose the cyber capabilities of the state posing the threat, if the threat of force is unjustified, then any future legitimate actions may not find support amongst the international community, or the Security Council for that matter. It will also become an easy suspect for any cyber attacks that may come to be carried out around the world. A cursory look at the scenario may present a negative answer, but a closer look and a more refined analysis may present a different view on the question.

B. Cyber Defence Shields: Necessary and Proportionate?

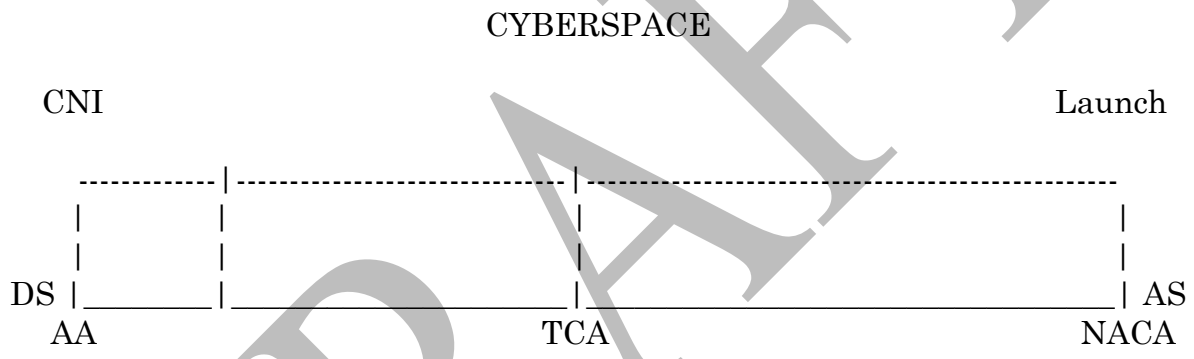

Fig 1: Perceived attack in cyber space showing 'interception points' in *AS*, *TCA*, and *NACA* from right to left.

The above diagram is used to help theorise the point at which an automated cyber response may take place and if such a response would fall within the cardinal requirements of necessity and proportionality. Several primary observations need to be made noting Haussler's distinction between (artillery) rockets and missiles (whether guided or otherwise),[153] and a cyber attack, where the target is a whole network as opposed to a specific target in real world terms.[154] Mid course or 'free flight phase' is denotes that a rocket or missile is not, or is no longer, guided at some point during its flight.[155] There are also technological constraints as to when interception may take place in different realms. Successful interception is more likely during the free flight or re-entry

[153] Francis Grimal, 'Missile Defence Shields: Automated and Anticipatory Self-Defence?' (2014) 19(2) *Journal of Conflict & Security Law* 317-339.

[154] This aspect of the cyber realm is introduced into the argument here, as there is a real difference in terms of time and space between the two sets of warzones/battlefields compared, *viz*., geographical landscape plus atmosphere, and cyberspace/ cyberscape.

[155] Grimal (n 153). 'Free flight phase' was normally used specifically for ballistic missiles, namely, in order to differentiate between the boost, free flight, and re-entry phases.

phases of ballistic missiles due to the time factor and speed.[156] Whereas the same cannot be said about a cyber attack in cyber space, as the free flight or re-entry phases are absent in cyber space, and also there are no equivalents that could be possibly analogised.

a) **NACA No Actual Armed Attack:** No necessity and proportionality—any action from **DS (Defending State)** will be unlawful.

In this context, the conclusion adopted, is that since a state has yet to suffer a cyber operation amounting to an actual 'armed attack', it is difficult to content that the necessity and proportionality elements have been satisfactorily met. The responding CDS / state arguably would not be fulfilling the 'last resort' requirement of necessity since it would have other options at its disposal. Similarly, since proportionality is intrinsically predicated on a response being proportionate to the defensive necessity of abating or repelling an attack, the fact that no such attack has occurred, renders the proportionality requirement somewhat moot.

b) **TCA (Threatened Cyber Attack):** Automated interception is unlawful unless the state can claim an action falling within anticipatory self-defence. Attribution is problematic against a TCA which, may allow defensive blocking falling short of a forceful response in the absence of the necessity and proportionality criteria being met.

A CDS / state acting against a threatened cyber attack will potentially struggle to meet the necessity and proportionality requirements. Although the CDS / state is potentially 'conscious' of threatened cyber attack, it may be impossible to discern or attribute the source of that attack. Consequently, this lack of attribution renders it difficult or perhaps nigh impossible to satisfy either of the necessity and proportionality elements at least in the practical sense. While the state may be in 'survival, last resort mode', it cannot practically resort to anything unless and until it has identified its attacker otherwise, the response will be disproportionate—the 'interception' may end up attacking innocent states.

c) **Actual Armed Attack:** *The closest one gets to impact the more likely automated is lawful – presumably attribution more likely detectable.* More probable. Necessity and proportionality possible providing attribution is possible.

This scenario is more easily reconciled within the traditional self-defence paradigm. Providing a CDS / state can pinpoint its attacker, the response is theoretically (at least) capable of meeting the necessity and proportionality threshold requirements. The cyber defence shield can presumably discern its attacker and then ascertain whether it has other options at its disposal and

[156] Grimal (n 153).

indeed (again, theoretically) calibrate its response accordingly to abating or repelling a further attack.

V. CONCLUSION

Much of the difficulty with applying the *jus ad bellum* to cyber-attacks concerns attribution (notably, because using proxy servers to hide the original IP address can comparatively easily mask the author of a cyber-attack).[157] Nazario observes that in the above cases of Estonia and Georgia (and a few other cases), from a political perspective classic right-wing sentiments are apparently behind the attacks, and it is noticeable that attackers use DDoS attacks to express support for an official government position, either against external or internal foes, besides also causing the victim some punitive damage to register their dissent with the victim's actions.[158]

In Nazario's view attribution continues to 'be a significant challenge in this problem space when retaliatory measures are considered,' and the attacks can spiral into significant diplomatic incidents if great care is not taken.[159] In both cases identified, circumstantial evidence appear to point in the direction of non-state players carrying out the cyber attacks, strongly backed by a state player, sharing a common agenda, but yet the same may not be sufficient to launch any offensive against the state player.

Maogoto's view, technological advances in warfare have 'overtaken' the international legal framework, which, he asserts, is now ill-equipped to regulate both aggressive cyber operations and attacks in/from outer space. For example, in relation to the right of self-defence, Maogoto argues (17) that the injury suffered by a state by virtue of intrusions into 'the digital commons' will likely not be sufficient to trigger Article 51: an attack that merely corrupts or 'annoys' rather than destroys would, in his view, fall short of a qualitatively grave use of force triggering a defensive response.[160] The fear is that this may leave states with no defensive recourse in response to technological threats.[161]

By way of overall conclusion, the authors of this article take the view that a) while automated and anticipatory self-defence is certainly strategically desirable, advancing such a legal argument remains problematic—at best an

---

[157] See P W Singer and Allan Friedman, *Cybersecurity and Cyberwar: What Everyone Needs to Know* (Oxford University Press, 2014) 75.

[158] Nazario (n 7) 172. See also Dorothy Denning, 'Activism, Hactivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy' in John Arquilla and David Ronfeldt (eds) *Networks and Netwars* (RAND, 2001) 239-288. The author presents a broad definition of cyberterrorism in the following terms: '…politically motivated hacking operations intended to cause grave harm such as loss of life or severe economic damage'. This definition almost touches upon most aspects under discussion in the instances under discussion, but whether the cyber attacks in question can be branded as 'cyber terrorism' is moot.

[159] Nazario (n 7) 174.

[160] This is the prevailing view in the literature. See, for example, Harrison Dinniss (n 1) 81. However, for a contrary position, see Nicholas Tsagourias, 'Cyber-Attacks, Self-Defence and the Problem of Attribution' (2012) 17 *Journal of Conflict and Security Law* 229, particularly at 231-2.

[161] See, for example, Matthew C Waxman, 'Self-Defensive Force against Cyber Attacks: Legal, Strategic and Political Dimensions' (2013) 89 *International Legal Studies* 109 (echoing this concern).

automated response in self-defence would have to be of a blocking nature,[162] and b) a theoretical/conceptual stretching of the temporal considerations is again desirable but an overly elastic interpretation is very much open to abuse. Unlike its IHL counterpart in the form of the Tallinn Manual, the *jus ad bellum* as this article as this article as sought to suggest it remains an area needing further clarity.

[162] We are grateful to Dr Duncan Hodges for offering his views on the issue.