Minds & Machines (2017) 27:387–392
DOI 10.1007/s11023-017-9446-1

Deterrence by Norms to Stop Interstate Cyber Attacks

Mariarosaria Taddeo[1,2]

Received: 9 September 2017 / Accepted: 9 September 2017 / Published online: 19 September 2017

In April 2017, the foreign ministers of the G7 countries approved a 'Declaration on Responsible States Behaviour in Cyberspace' (G7 Declaration 2017). The Declaration addresses a mounting concern about international stability and the security of our societies after the fast-pace escalation of cyber attacks occurred during the past decade. In the opening statement, the G7 ministers stress their concern

> […] about the risk of escalation and retaliation in cyberspace […]. Such activities could have a destabilizing effect on international peace and security. We stress that the risk of interstate conflict as a result of ICT incidents has emerged as a pressing issue for consideration. […], (G7 Declaration 2017, 1).

Paradoxically, state actors often play a central role in the escalation of cyber attacks. State-run cyber attacks have been launched for espionage and sabotage purposes since 2003. Well-known examples include Titan Rain (2003), the Russian attack against Estonia (2006) and Georgia (2008), Red October targeting mostly Russia and Eastern European Countries (2007), Stuxnet and Operation Olympic Game against Iran (2006–2012). In 2016, a new wave of state-run (or state-sponsored) cyber attacks ranged from the Russian cyber attack against Ukraine power plant,[1] to the Chinese and Russian infiltrations US Federal Offices,[2] to the Shamoon/Greenbag cyber-attacks on government infrastructures in Saudi Arabia.[3]

[1] https://www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/.

[2] https://www.nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html?_r=0.

[3] https://www.symantec.com/connect/blogs/greenbug-cyberespionage-group-targeting-middle-east-possible-links-shamoon.

✉ Mariarosaria Taddeo
mariarosaria.taddeo@oii.ox.ac.uk

[1] Oxford Internet Institute, University of Oxford, Oxford, UK

[2] Alan Turing Institute, London, UK

This trend will continue. The relatively low entry-cost and the high chances of success mean that states will keep developing, relying on, and deploying cyber attacks. At the same time, the ever more likely AI leap of cyber capabilities (Cath et al. 2017)—the use of AI and Machine Learning techniques for cyber offence and defence—indicates that cyber attacks will escalate in frequency, impact, and sophistication.

Historically, escalation of interstate conflicts has been arrested using offensive or political strategies, sometimes in combination. Both have been deployed in cyberspace. The first failed; the second needs to be consolidated and enforced (Taddeo and Glorioso 2016a, b).

1 The Offensive Strategy

Escalation follows from the nature of cyber attacks and the dynamics of cyberspace (Floridi and Taddeo 2014; Taddeo 2014, 2016, 2017). Non-kinetic cyber attacks—aggressive attacks that do not cause destruction or casualties, e.g. deploy zero-day exploits or DDoS attacks—cost little in terms of resources and risks to the attackers, while having high chances to be successful. At the same time, cyber defence is porous by its own nature (Morgan 2012): every system has mistakes or bugs in the program (vulnerabilities), identifying and exploiting them is just a matter of time, means, and determination. This makes even the most sophisticate cyber defence mechanisms ephemeral and, thus, limits their potential to deter new attacks. And even when successful, cyber defence does not lead to strategic advantages, insofar as dismounting a cyber attack may bring tactical success, but very rarely leads to the ultimate defeating of an adversary (Taddeo 2017). This creates an environment of persistent offence (Harknett and Goldman 2016), where attacking is tactically and strategically more advantageous than defending.

In this scenario, state actors make policy decisions to protect their abilities to launch cyber attacks. 'Strategic ambiguity' is one of such decisions. According to this policy, states decide neither to define and nor inform the international community about their *red lines*—thresholds that once crossed would trigger state response—for non-kinetic cyber attacks (Taddeo 2011).

Strategic ambiguity has often been presented as a way to confuse the opponents about the consequences of their cyber attacks. As the US National Intelligence Officer for Cyber Issues officer put it:

> Currently most countries, including ours, don't want to be incredibly specific about the red lines for two reasons: You don't want to invite people to do anything they want below that red line thinking they'll be able to do it with impunity, and secondly, you don't want to back yourself into a strategic corner where you have to respond if they do something above that red line or else lose credibility in a geopolitical sense.[4]

By fostering ambiguity, state actors also leave open for themselves a wider room for manoeuvring. Strategic ambiguity allows state actors to deploy cyber attacks for

[4] http://www.c4isrnet.com/articles/cyber-red-lines-ambiguous-by-necessity.

military, espionage, sabotage, and surveillance purposes without being constrained by their own policies or international red lines. This makes ambiguity a dangerous choice, one that is strategically risky and politically misleading.

The risks come with the cascade effect following the absence of clear thresholds for cyber attacks. The lack of thresholds facilitates a proliferation of offensive strategies. This, in turn, favours an international cyber arms race and the weaponization of cyberspace, which ultimately spurs the escalation of cyber attacks. In parallel, while seeking to maintain uncertainty about red lines to deter prospective cyber attacks, it actually ends up leaving unbounded (state and non-state run) non-kinetic cyber attacks, which are indeed the great majority of cyber attacks.

Clearly, short of an ultimate victory, offensive strategy leads to policy hazards that fuel, rather than arresting, escalation of interstate cyber attacks. Cyber attacks would be deterred more effectively by a regime of international norms that makes attacks politically costly to the point of being disadvantageous for the state actors who launch them.

2 The Political Strategy

Over the past twenty years, the UN, the Organisation for Cyber Security and Co-operation in Europe (OSCE), and the ASEAN Regional Forum (ARF) and several national governments (G7 and G20) have convened consensus to define such a regime. The time has now come to strengthen and implement it.

The G7 Declaration is the latest of a series of successful transnational initiatives made in this direction before the recent failure of the UN Group of Government Experts (UN GGE) on 'Developments in the field of information and telecommunications in the context of international security'.[5]

Like the UN GGE recommendations, the G7 Declaration identifies two main instruments: confidence building measures (CBMs) and voluntary norms. CBMs foster trust and transparency among states. In doing so, they favour co-operations and measures to limit the risk of escalation. CBMs range from establishing contact points, shared definitions of cyber-related phenomena, and communication channels to reduce the risk of misperception, and foster multi-stakeholder approach.

Voluntary norms identify non-binding principles that shape state conduct in cyberspace. De facto, voluntary norms identify red lines for state run non-kinetic cyber attacks and, thus, fill the void created by strategic ambiguity. They stress that states should not target critical infrastructures and critical information infrastructures of the opponent (norms 6, 8, and 11 of the G7 Declaration); should avoid using cyber attacks to violate intellectual property (norm 12 of the G7 Declaration); and remark the responsibility of state actors to disclose cyber vulnerabilities (norms 9 and 10 of the G7 Declaration).

[5] https://www.justsecurity.org/42768/international-cyber-law-politicized-gges-failure-advance-cyber-norms/.

> Norms 9 and 10 are particularly significant as they stress that
>
> 9. […] States should seek to prevent the proliferation of malicious ICT tools and techniques and the use of harmful hidden functions;
>
> 10. States should encourage responsible reporting of ICT vulnerabilities and share associated information on available remedies to such vulnerabilities to limit and possibly eliminate potential threats to ICTs and ICT-dependent infrastructure.

These norms tackle one of the key mechanisms of the cyber arm race: state actors acquiring cyber vulnerabilities with the aim of exploiting rather than disclosing or patching them. Norms 9 and 10 have been widely mentioned in the aftermath of the recent WannaCry attack, which ran on a vulnerability (EternalBlue) identified and not disclosed by the UN National Security Agency.[6] If respected, these norms would have helped to avert the attack. Interestingly, despite being one of the signatories of the G7 Declaration (as well as being represented in the 2015 UN GGE that first drafted these principles), the US did not face any measures or sanctions following the violation of these norms. This unveils the main problem with existing political strategies to stop escalation of cyber conflicts: the lack of any binding, coercive strength.

The voluntary, non-binding nature of these norms transforms them—and the entire G7 Declaration—in a formal exercise, devoid of any political strength, and thus unable to reach its goal. A step forward is necessary to overcome this stalemate.

The time has arrived for international consensus to be turned into multi-lateral agreements, and to transform voluntary norms into binding ones. This will provide the foundation for an international regime of norms delineating permissible and non-permissible state actions in cyberspace. Without this regime, cyber attacks will continue to be the elective choice of state actors and will contribute to fuelling cyber arm race, making cyber stability a chimera.

Once defined and agreed upon, this regime will have to be enforced. The enforcement requires an independent authority able to exert coercive power and impose sanctions. This authority cannot (and should not) be the result of a multi-stakeholder or a neutral, private-led initiative. This would impose too heavy responsibilities on the private sector and create an authority too weak to bear the political pressure resulting from ensuring state compliance with an international cyber regime. In the same vein, national internal policies for states to verify their own compliance with the regime would be just a different way to endorse voluntary norms.

Enforcing this regime requires an authority able to (1) convene agreement about international norms, (2) verify states compliance with the norms at an international level, (3) launch investigations into suspected state-run (or state-sponsored) cyber attacks to ascertain attribution, (4) expose breaches of the norms and the sources of illegitimate cyber attacks, and (5) impose adequate sanctions and punishments. Achieving (1)–(5) necessitates the coordination of intelligence, political, and

[6] https://www.forbes.com/sites/leemathews/2017/05/17/wannacry-ransomware-wasnt-the-first-malware-using-stolen-nsa-exploit/#56c7ba594bd8.

diplomatic capabilities, and extremely advanced technical skills, as well as the authority and apparatus to enforce sanctions and punishment. (1)–(5) define a politically-loaded mandate for an authority that will have a deep impact on international relations and geo-political equilibriums.

The mandate resonates perfectly well with Article 26 of the UN Charter, which defines the mission of the Security Council:

> […] to promote the establishment and maintenance of international peace and security with the least diversion for armaments of the world's human and economic resources, the Security Council shall be responsible for formulating, with the assistance of the Military Staff Committee […] plans for the establishment of a system for the regulation of armaments.[7]

Indeed, the UN Security Council has the necessary resources, the political, and coercive power to deliver successfully (1)–(5). The time has come to embrace this power to consolidate and enforce an international regime of norms to deter cyber attacks and limit cyber arm race, while fostering peace. Problems, mistakes, and even failures—like the failure of the UN GGE to agree on norms, rules, and principles for responsible state's conduct in cyberspace—are to be expected but they must not hinder the process. The alternative is a militarized cyberspace threatening, rather than fostering, the flourishing of our societies.

In a vicious cycle, cyber attacks and cyber arm race feed one each other. Together, they pose serious threats to the stability of cyberspace and, in turn, to the security and the peace of information societies. Where offensive strategies have failed to break this cycle, political strategies must succeed.

References

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the 'Good Society': The US, EU, and UK approach. *Science and Engineering Ethics*. doi:10.1007/s11948-017-9901-7.

Floridi, L, & Taddeo, M., (Eds.) (2014). *The ethics of information warfare*. Law, governance and technology series, Vol 14. Heidelberg: Springer.

G7 Declaration. (2017). G7 Declaration on responsible state behavior in cyberspace. Lucca. http://www.mofa.go.jp/files/000246367.pdf..

Harknett, R. J., & Goldman, E. O. (2016). The search for cyber fundamental. *Journal of Information Warfare, 15*(2), 81–88.

Morgan, P. M. (2012). The state of deterrence in international politics today. *Contemporary Security Policy, 33*(1), 85–107. doi:10.1080/13523260.2012.659589.

Taddeo, M. (2011). Information warfare: A philosophical perspective. *Philosophy and Technology, 25*(1), 105–120. doi:10.1007/s13347-011-0040-9.

Taddeo, M. (2014). Just information warfare. *Topoi*. doi:10.1007/s11245-014-9245-8.

[7] http://www.un.org/en/sections/un-charter/chapter-v/.

Taddeo, M. (2016). On the risks of relying on analogies to understand cyber conflicts. *Minds and Machines, 26*(4), 317–321. doi:10.1007/s11023-016-9408-z.

Taddeo, M. (2017). Cyber conflicts and political power in information societies. *Minds and Machines, 27*(2), 265–268. doi:10.1007/s11023-017-9436-3.

Taddeo, M, & Glorioso, L., (Eds.) (2016a). *Ethics and policies for cyber operations*. Philosophical studies, Springer.

Taddeo, M., & Glorioso, L., (Eds.) (2016b). Regulating cyber conflicts and shaping information societies. In *Ethics and policies for cyber operations*. Philosophical studies series. Berlin: Springer.