CCDCOE
NATO COOPERATIVE
CYBER DEFENCE
CENTRE OF EXCELLENCE

Cyber vigilantism in support of Ukraine: a legal analysis

Ann Väljataga

March 2022

Introduction

Reportedly, around 40 000 people from 52 countries have taken President Volodymyr Zelensky up on his offer to 'join the defence of Ukraine, Europe and the world,' and enrolled in the International Legion of Territorial Defence of Ukraine. An additional 300,000 (and counting) have responded to the Tweet from Ukraine's Minister of Digital Transformation, Mykhailo Fedorov, which called for IT professionals from around the world to join Ukraine's IT army. There are also independent cyber vigilantes organising themselves in a more nebulous fashion and thus escaping any association with state agencies. Shortly before the invasion, a Belarusian-based hacktivist group called the Belarusian Cyber Partisans encrypted 'the bulk of the servers, databases and workstations' of Belarusian Railway to 'slow down the transfer' of Russian troops and succeeded in stopping railway traffic in Minsk, Orsha and Osipovichi. Should Russia not pull back, Anonymous has threatened to take industrial control systems hostage. AgainstTheWest's (ATW) Russian-oriented prong announced on Twitter that it had breached the systems of the Russian Space Forces, the Ministry of Transport of Russia and Russia Air. Initiatives of active cyber resistance have been met with both praise and reprimand. As much as they are clever and innovative in taking a bottom-up grassroots approach to countering injustice and violence, they are also legally ambiguous and disposed to more serious consequences than initially planned. Robert M. Lee, CEO of an industrial cyber security company Dragos, who led the investigation of the cyber attacks against Ukrainian power grids in 2015, put it bluntly by saying that, '[a]nyone not working on behalf of a government having serious conversations about 'hacking back' or launching cyber attacks against Russia please understand – respectfully – you're an idiot and only going to make matters worse'. Such assessments aside, this paper seeks to explain what is at stake from the international law perspective and analyses three specific factors that have a particular effect on the legal evaluation of hacktivist cyber operations in times of armed conflict.

1. The position of the hacktivist

Individuals who are not connected to national defence structures through their professional or voluntary activities quite clearly have more freedom to express their support to a foreign state entangled in an armed confrontation. Because their active support might be seen as a form of state participation, the options available to military personnel, members of voluntary national defence organisations and state officials are understandably more limited. Many states have expressed their positions concerning their citizens and residents going to Ukraine to support the fight against the Russian invasion, but these statements have not touched on the possibility of contributing through cyber means.

Estonia has recognised that its citizens and residents who are not part of the national defence institutions are legally allowed to join Ukrainian defence forces; the recognition does not, however, equate to endorsement and does not extend to separatist groups, or the Russian army, nor does it translate into express approval. To decriminalise fighting on Ukraine's side, Latvia has introduced amendments to its national legislation, the Prime Minister of Czech Republic has publicly stated that foreign fighters joining Ukraine's forces would not be prosecuted for fight-related charges. The UK has sent out mixed signals ranging from approval to explicit prohibition. However, its national legislation (Foreign Enlistment Act of 1870 and possibly also the Counter-Terrorism and Border Security Act of 2019) prohibits participating in armed conflicts abroad. Germany has criminalised recruitment but not participation,

Denmark and Canada have condoned fighting in the ranks of the Ukrainian army, and Japan has made clear that fighting in a foreign armed conflict is and remains punishable under its Penal Code. While the majority of the approving statements concern enlisting in the International Legion of Territorial Defence of Ukraine and no other groups, Germany, for example, has made no such distinction. The establishment of the Legion was announced by the Ukrainian Minister of Foreign Affairs and is based on a presidential decree from 2016 that allows non-Ukrainians to join the armed forces of Ukraine. Despite this, according to the Russian Ministry of Defence foreign fighters joining the Legion will not be viewed as combatants under international law, meaning that they are not entitled to the status of prisoner of war (POW).

2. The structure, organisation and government affiliations of the group

Unlike the Legion, the IT Army is unlikely to be viewed as an organised armed group for the purpose of determining combatant status, and so many of the questions relating to combatancy and the different forms of civilian participation matter greatly. For instance, although the privilege of being treated as a POW might not hold much significance for a cyber fighter, other aspects such as becoming a legitimate military target[1] and having limited legal immunity from criminal prosecution [2] might become very important.

Although international in its composition, the IT Army was created and is to an extent coordinated by the Ukrainian government, most directly by the Ministry of Digital Transformation, which puts it under the effective control of Ukraine. However, despite a clear relationship with a belligerent state, members of the IT Army are unlikely to meet the criteria set out for status as combatants as the vast majority do not belong to the Ukrainian armed or irregular forces and the IT Army is not officially recognised as part of the Ukrainian defence apparatus. Therefore, the IT Army could be best described as a cross between state-sponsored hacking and decentralised hacktivism.

Depending on the nature of the operations it undertakes and how it is governed and organised, the volunteers in the IT Army can be legally categorised as either civilians indirectly supporting hostilities, civilians directly participating in hostilities or potentially *levée en masse*. The latter option, however, entails elements that make applying it to contemporary hacktivism problematic (see 4.c). If the likes of Anonymous or NB65 operate in an obfuscated non-hierarchical manner without being instructed by or answerable to any state organ, its members might qualify as civilians directly participating in hostilities, civilians indirectly supporting hostilities or simply criminals. The distinction is of importance since the first categorisation turns the involved civilians into legitimate military targets under IHL, while the latter subjects them to peacetime law enforcement procedures.

A) Direct participation in hostilities

Geographical distance from the battlespace does not necessarily rule out direct participation. For instance, the Israeli Supreme Court has in its *Targeted Killings* opinion found that a person, despite the considerable distance from the battlefield, directly takes part in hostilities if they operate or supervise the operation of or service a system.[3] While there is no case law or *opinio juris* which examines civilian participation in the cyber domain, given the advances in modern warfare the prerequisite of geographical proximity might well have become redundant. [4] Regardless of the location of the act, three elements determine direct participation in hostilities (DPH):

- the 'threshold of harm' must be met, which extends to any consequence adversely affecting the military operations or military capacity of a party to the conflict;

- a direct causal link between the act and the harm resulting or likely to result either from that act or from a coordinated military operation of which that act is an integral part (direct causation); and
- the intent in committing the act to directly cause the required threshold of harm in support of a party to the conflict and the detriment of another (belligerent nexus).[5]

At first glance, the criteria might seem restrictive, but in practice, DPH has been interpreted to include collecting intelligence on the armed forces,[6] the supply and transportation of weapons (particularly in cases where there is an established geographical link to the battlefield)[7] or personnel[8] and engaging in sabotage and disruption of the enemy's communications.[9] Operations below the threshold of a cyber attack will satisfy the criteria as long as they have a negative effect on the enemy's military activities.[10] There are thus no guarantees that a commitment to avoid targeting critical infrastructure or otherwise causing widespread direct damage will ensure the continuity of civilian status. States' military manuals also seem to favour a case-by-case approach and draw no definitive lines.[11]

If an individual engaged in a cyber activity against a belligerent state falls within the definition of a civilian DPH, they become a legitimate military target.[12] This means that, provided they systematically participate in hostilities, military force can be used against them.[13] Retaliatory or preventive use of force against an individual only 'sporadically' participating in hostilities is deemed unlawful. Participation is considered systematic and constituent throughout the phases of preparation, target identification, active operation and after-action assessments. A hacktivist directly participating in hostilities is targetable during the preparative stages including collecting intelligence on targets such as Belarusian Railway, Russian scientific research satellites or state-run TV transmission platforms. Likewise, they remain a legitimate object of military attacks while obtaining access to the target system, running the malware and throughout any after-action assessment of the results and the need for a repeat attack. While it might not always be technically possible or, for that matter, strategically advantageous for a state to go after hacktivists, international law sets no rigid boundaries and digital participation in hostilities should be taken as seriously as its kinetic equivalents. Although international law generally favours the presumption against direct participation, the continued validity of such a presumption has been questioned.[14] For the mercenaries and foreign volunteers and for those involved in transporting weapons to Ukraine, Russia has signalled that it will define legitimate military targets and direct participation broadly.

In brief, for an individual planning on engaging in political hacktivism amidst the armed conflict between Russia and Ukraine, the most important factors to ponder would be the nature and effects of the planned operation and their personal status, nationality and geographical location. Since civilians supporting Ukraine by carrying out cyber operations designed to have the gravest impact – and 'taking hostage' industrial control systems is likely to qualify – may, pursuant to international humanitarian law, invoke a cross-border military response in any of the operational domains, refraining from such operations will be more likely to ensure individual safety and avoid escalation. Although many such declarations have posed no valid threat grounded in international law, Russia's space research agency Roscosmos has taken the line that a cyber operation that caused its reconnaissance satellites to lose contact with the ground station was an 'act of war'. The affected satellites arguably carried out scientific research and had filled no dual-use or military functions whatsoever. Examples like this once again indicate that Russia is currently willing to interpret any cyber operation in the most aggressive and escalatory terms.

B) Indirect participation in hostilities

Indirect participation includes the means of supporting a state party's war that do not meet the three criteria. Historically, this has included providing logistical assistance, financial support, food, awareness-raising and propaganda.[15] In the cyber domain, it is likely to entail passive defence of Ukraine's critical networks and a large share of operations directed at awareness-raising or halting the enemy's disinformation campaigns. Intelligence, data exposure and DDoS attacks should be analysed individually, taking into account any causal link to definitive military harm. Take, for instance, the case where the IT Army broke into the Kremlin's phone registry and disclosed the numbers of all employees while urging all Ukrainians and others to call and record the conversations so that the recordings could later be used as evidence in judicial proceedings. As it entailed unauthorised access to a system, the operation constituted a criminal offence but failed to contribute to military harm to an extent amounting to direct participation. Consequently, while the individuals who took part in the operation can be charged under criminal law, they retain their status as civilians and will not become legitimate military targets.

This will also apply in cases where such operations are conducted by groups not affiliated with nation-states. Perhaps the best example of this was from the Russian Federal Service for Supervision of Communications, Information Technology and Mass Media ('Komnadzor'). On 10 March, Anonymous claimed that it had breached the database of the agency, leaking over 360,000 files in the process. The documents were published on the Distributed Denial of Secrets website. The long-term effects of a leak of that scale are yet to be seen; however, it is far-fetched to expect them to amount to a direct contribution to military harm. On 26 February, at the beginning of the invasion, Anonymous claimed that it had disrupted the transmission of state-run Russian TV channels and reprogrammed the stream of images from the war in Ukraine and messages opposing it. The disruptions were reversible, temporary and intended to raise awareness among the Russian public of the atrocities being committed. Since mere spreading of information has previously been excluded from the scope of DPH[16] and there is no evident causal link to military harm, the perpetrators did not risk losing their civilian status, but they nevertheless clearly committed a series of cyber offences criminalised by the Russian Penal Code[17] and by most other nations that have laws on cybercrime.

C) Levée en masse

In an international armed conflict, inhabitants of an unoccupied territory who engage in cyber operations as part of a *levée en masse* enjoy combatant immunity and POW status. A *levée en masse* is composed of inhabitants of the territory over which the war is waged but which is not yet occupied by foreign forces.[18] The concept was originally introduced so that people spontaneously standing up against an invasion without having the time to formally organise into combat units could be granted the responsibilities and privileges of combatants (POW status and limited legal immunity). [19] Hacktivist groups that are composed of members from around the world and have no intention to obtain a structure reminiscent of combat units are unlikely to meet the criteria. As to the Ukrainian IT Army, its level of organisation and subordination to the Ukrainian government seems a degree too high for it to be viewed as a *levée en masse*. Then again, the requirement of being formed of 'inhabitants' does not apply in the cyber domain. Also, since *levée en masse* presumes a spontaneous uprising of the general population, fitting groups that are united by specific skillset under the notion might prove equally problematic. Finally, members of *levée en masse* are required to carry their arms openly. When forcefully moulded to accommodate hacktivist groups, the requirement can easily subside into something as nonsensical as an obligation not to actively hide one's personal computer. Applying a condition as deeply rooted in an entirely different technological environment to

present-day hacktivism would hence mark a stretch too far from its original idea.

3. A cyber duty of due diligence as the basis for state responsibility

While the operating model of hacktivist groups like Anonymous, ATW or NB65 eschews government affiliations, states might nevertheless be held accountable for the activities of groups operating within their jurisdiction. Should it prove strategically beneficial to the target state, it could seek to attribute hacktivist operations to a state by referring to a breach of a due diligence obligation. In essence, this presumes that, although the state from which the attacks were launched was capable of stopping them, it wilfully or negligently failed to do so. This avenue was sometimes hinted at in the aftermath of the cyber attacks against Estonia in 2007 and Georgia in 2008, which seemingly originated from a group of ambiguously organised Russian civilians not supervised by any state agency. On both occasions, Russian intelligence and law enforcement displayed a reluctance to exercise their power to investigate or prosecute the wave of cyber hostilities emanating from Russian territory.

Although official public attribution has become more common since then, the breach of due diligence obligation has not yet been used as the primary basis for evoking state responsibility for cyber activities. The war in Ukraine will probably not bring a paradigm shift in this, especially since the cyber attacks against Russia and its client state Belarus seem to be truly diffuse in their geographical origins. In principle, however, having established that another state has failed to exercise due diligence in preventing the harm and investigating the cyber attacks being carried out in its territory, the target state can respond to a breach of sovereignty or prohibited intervention by, for example, imposing sanctions, damaging or disrupting the systems from where the attacks are launched, blocking internet traffic from certain countries, expelling diplomats or some other unfriendly or unlawful action not amounting to the use of force. Stemming from the principle of necessity, any response can only serve the objective of stopping the ongoing cyber operation and not that of deterrence, punishment or prevention.[20] Regardless of the many practical and procedural obstacles that a state might face when seeking recourse against another state that has allowed its territory to be used for wide-scale cyber hostilities, international law does not exclude the possibility of an aggressive interstate response to civilian cyber activities. Moreover, in today's tense political climate, such scenarios may not only have credibility, but also the potential to cause a spillover of the ongoing conflict. Any speculation about the actual capacity of a morally bankrupt and economically ravaged war-torn pariah state to go from mere naming and shaming to applying effective countermeasures is outside the scope of this paper.

Treading on a dangerous ground: a conclusion

In the Ukraine-Russia war, civilian cyber volunteers have become the most vocal actors on the cyber front, partially because hacktivism is innately loud and partially because whatever happens on a state-to-state level, where the stakes are considerably higher is stealthy and silent. However, it seems like the Ukrainian IT Army, which is more strategic and precautious in its operations and the various hacktivist groups seem to be treading the thin lines drawn by international law. They make the most of information leaks, data sharing and web defacements, maximise the effects of temporary disruptions and quotidian nuisance without becoming irrevocably involved in the armed conflict. The instances that seem the closest to qualifying as direct participation have been committed against Belarus, a state that despite its obvious leanings is not (yet) at war. The attacks directed at various Russian government websites and information systems or that have interfered with news and broadcasting services will, unless they somehow directly contribute to military harm, leave the perpetrators' civilian immunity intact.

Be it due to mere luck or skilful deliberation, the lack of serious casualties or legal consequences cannot be taken for granted, particularly when considering Russia's apparently liberal use of the term 'act of war' and the swift changes in Russian criminal law.

To avoid the conflict spilling over because of a hack gone wrong, before signing up for any activity inspired by the desire to help a state that has fallen victim to an unjustifiable invasion, an individual should bear in mind their position in relation to any state organ and that of the particular hacktivist group they plan to join. The most important factor that decides the legality and overall safety is still the character of the activity; while information leaks and anti-propaganda hacks do not turn a civilian into a lawful or unlawful combatant, they do jeopardise ongoing government intelligence operations. On the contrary, hacktivist operations of potentially uncontrollable or indiscriminate effects such as open-source supply chain attacks or anything targeting critical infrastructure are never worth the risks – military, humanitarian or criminal – that they pose.

---

[1] Additional Protocol I to the Geneva Conventions, Article 43.
[2] Ibid.
[3] Supreme Court of Israel, Public Committee against Torture in Israel v. Government of Israel, Case No. HCJ 769/02, 13 December 2006, para 35, http://elyon1.court.gov.il/files_eng/02/690/007/A34/02007690.a34.pdf.
[4] See also ICRC, Third Expert Meeting on the Notion of Direct Participation in Hostilities Geneva, 23 – 25 October 2005 , Summary Report Co-organised by the International Committee of the Red Cross and the TMC Asser Institute. https://www.icrc.org/en/doc/assets/files/other/2005-09-report-dph-2005-icrc.pdf.
[5] ICRC Interpretive Guide on the Notion of Direct Participation in Hostilities under IHL (DPHIG), pp 49-64.
[6] ICTY, The Prosecutor v. Pavle Strugar, Case No. IT-01-42-A, Appeal Judgement, 17 July 2008, para. 177.
[7] Ibid, 177; *United States of America v. Salim Ahmed Hamdan*, U.S. Military Commission, 19 December 2007, p. 6.
[8] *Supra* note 6.
[9] Prosecutor General to the German Federal Court of Justice, Fuel Tankers case, Case No. 3 BJs 6/10-4, 16 April 2010, pp 59-63, https://ihl-databases.icrc.org/customary-ihl/eng/docs/v2_cou_de_rule14. .
[10] Schmitt, M.N. ed., 2017. Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (hereafter Tallinn Manual 2.0) Cambridge University Press, Rule 97(6), p.430.
[11] *Supra* note 5.
[12] Article 51(3) of Additional Protocol I to Geneva Conventions and Article 13(3) of Additional Protocol II.
[13] Ibid.
[14] See eg Schmitt, M.N., 2011. The interpretive guidance on the notion of direct participation in hostilities: a critical analysis. In *Essays on Law and War at the Fault Lines*, TMC Asser Press, p. 24.
[15] *Targeted Killings*, *supra* note 4.
[16] Targeted Killings, *supra* note 4, para 39.
[17] Penal Code of the Russian Federation, see Articles 272 – 274, 207.3 (207.3 dissemination of 'fake news'- recently amended)
[18] Additional Protocol III to the Geneva Conventions , art 4A(6).
[19] Tallinn Manual 2.0, *supra* note 11, Rule 88.
[20] Draft Articles on Responsibility of States for Internationally Wrongful Acts (DARS), Arts 49-53