Review Protocol: Investigating Cyber–warfare and Compliance with International Humanitarian Law (v1.0)

Lead Researcher: SN 20180626

21st February 2021

1 Summary

The Fourth Geneva Convention was ratified in 1949 [1], long before the advent of modern computers and the internet. As offensive cyber action becomes increasingly prominent in the new "fifth dimension" of warfare [2], with rapid capacity–building for both offence and defence occurring in many states [2], there is significant debate between stakeholders across the public, private and third sectors regarding whether Customary International Humanitarian Law (IHL) [3] is sufficient to address this new "wicked problem" [4].

This systematic review will be approached from a constructivist perspective, with its purpose being to establish an understanding of the policy issue of IHL compliance in cyber–warfare. The objective of this review is to provide policymakers with a high–level overview of the policy issue and efficacy of current interventions. To achieve this objective, salient research literature will be systematically identified, collected and critically assessed for inclusion in a qualitative synthesis, from which answers may be inferred to the review's research questions (See **SECTION 2**).

2 Research Question

Can Customary International Humanitarian Law (IHL) sufficiently address cyber–warfare [4]?

2.1 Sub–questions

- Can cyber–attacks meet the principles of distinction, proportionality, humanity and military necessity to justify use of force, as per Customary IHL [3]?
- Does offensive cyber action comply with Just War Theory [5]?
- Can cyber–weapons be adequately assessed by the regulatory instrument of Article 36 weapon reviews, as per Geneva Conventions Protocol I [6]?

3 Academic Databases

The review will collect studies from the following academic databases:

- **ACM Digital Library** – journal articles, conference proceedings, magazines and other literature concerning computer science and information and communication technologies.
- **IEEEXplore** – journal articles, conference proceedings, technical standards and other literature concerning computer science, electrical engineering, electronics and other related fields.
- **ProQuest** – multi–disciplinary database including journal articles, conference proceedings, trade publications and other literature.
- **Scopus** – multi–disciplinary database including journal articles, conference proceedings, trade publications, books and other literature.
- **Web of Science** – multi–disciplinary database including journal articles, books and conference proceedings.

4 Disciplines

- Political science
- Military studies
- Law
- International relations
- Peace and conflict studies
- Technology
- Public policy and administration

5 Inclusion and Exclusion

The review will include:

- All types of study design and methodologies, including but not limited to, quantitative, qualitative and mixed–method reports, literature reviews, surveys, case studies and technical reports.
- All articles concerning the issue of cyber–warfare conducted by state, state–sponsored and non–state actors in the context of IHL, including cyber–attacks, cyber–espionage, cyber–weapons and cyber–deterrence.

The review will exclude:

- All articles that have not been published in peer–reviewed academic journals.
- Grey literature, including trade journal and government publications.
- All articles concerning the issues of information warfare / disinformation / propaganda, hacktivism, lethal autonomous weapons systems (LAWS), cyber–security, cyber–crime and cyber–terrorism.

6 Search Strategy

The review will use the following query strings containing salient keywords to interrogate academic databases (See **SECTION 3**):

- `(cyber AND war*) AND (law AND ("international humanitarian" OR "armed conflict"))`
- `(cyber AND attack*) AND (principle* AND (distinction OR necessity OR proportionality OR humanity))`
- `(cyber AND offens* AND action) AND (jus* AND (theory OR tradition OR "ad bellum" OR "in bello"))`
- `(cyber AND weapon*) AND ("article 36" AND review)`

Searches will be limited to articles published between 2007–to–present. 2007 represented an epoch in cyber–warfare with significant cyber–attacks suffered by Estonia [7], allegedly perpetrated by the Russian security apparatus [7]. These events prompted the formation of the NATO Cooperative Cyber Defence Centre of Excellence in Tallinn, Estonia [8], and the publication of the Tallinn Manual on the International Law Applicable to Cyber Warfare [9] (See **SECTION 10.1**).

7 Literature Management

As I am writing in the LaTeX typesetting format, I am managing references for each stage of the systematic review in corresponding BibTeX databases, using JabRef as my reference manager and Git for version control. I am hosting the .tex source for this document and corresponding .bib databases on a public GitHub repository (See **APPENDIX**).

8 Selection of Studies

I am using the PRISMA methodology for systematic reviews [10], beginning with identifying and enumerating salient studies collected from selected academic databases (See **SECTION 3**), using query strings containing salient keywords (See **SECTION 6**) to interrogate the databases. After removing duplicate records, I will screen the articles using the inclusion criteria (See **SECTION 5**) and excluding those which do not meet it. Excluded articles will be kept in a separate .bib database. Articles will then be assessed for eligibility, with excluded articles kept in another .bib database (See **SECTION 10**). The included studies will be rigorously assessed in a qualitative synthesis in the final review (See **SECTION 9**).

9 Strategy for Data Synthesis

Articles included for qualitative synthesis will be coded to extract salient information, such as research characteristics, data collection methodologies and findings. This information will then be used to map the research as a preparatory step for the data synthesis, potentially providing context that may inform the interpretation of the included articles. As this systematic review aims to establish an understanding of the issue (See **SECTION 1**), I will utilise a configurative approach to synthesise the included articles, employing inductive reasoning to integrate and interpret them. This synthesis will aim to answer the research questions (See **SECTION 2**) by identifying commonalities and divergences in the qualitative research findings of the included articles.

10 Included Studies and Preliminary Results

Stage	**Total n**
Database search results	5861
Duplicates removed	4690
Screened (title-abs-key)	852
Excluded (title-abs-key)	3838
Assessed (full page)	62
Excluded (full page)	770
Included (qualitative)	49

Table 1: PRISMA Flow

Search results from the selected academic databases (See **SECTION 3**), interrogated using query strings (See **SECTION 6**) and database APIs where possible, returned 5861 results. After cleaning the reference database and removing duplicates, 4690 articles remained and were prioritised for screening at the title, abstract and keyword level (title-abs-key). To automate the screening process, I leveraged machine learning by installing and deploying the open–source Active learning for Systematic Reviews (ASReview) software (See **APPENDIX**) on a dedicated cloud–based virtual machine. Of the 4690 articles prioritised for screening, 852 met the criteria for full page assessment (See **SECTION 5**), with 3838 articles excluded. To automate the process for full page assessment, I again leveraged machine learning by cloning and deploying the open–source General Architecture for Text Engineering (GATE) GitHub repository (See **APPENDIX**) on another dedicated cloud–based virtual machine. Of the 852 articles prioritised for full page assessment, 62 met the criteria for inclusion in the evidence synthesis (See **SECTION 5**), with 770 articles excluded. Of the remaining 62 articles, 49 were prioritised for inclusion in the qualitative synthesis.

10.1 Description of Included Studies for Qualitative Synthesis

Country	Total n	Keyword	Total n	Year	Total n	Subject	Total n
United States	18	Cyberspace	5	2020	4	Social Sciences	43
United Kingdom	15	Tallinn Manual	5	2019	5	Engineering	12
Australia	3	Cyber	4	2018	4	Arts and Humanities	10
Canada	3	Cyber Operations	4	2017	4	Computer Science	10
Netherlands	3	Attribution	3	2016	6	Business	2
South Korea	2	Cyber Attack	3	2015	8	Decision Sciences	2
Austria	1	Cyber Security	3	2014	0	Psychology	2
Belgium	1	Cyber War	3	2013	7	Economics	1
Denmark	1	Ethics	3	2012	7	Environmental Science	1
Germany	1	International Humanitarian Law	3	2011	2		
Greece	1	International Law	3	2010	2		
Israel	1	Jus Ad Bellum	3				
Japan	1						
Norway	1						
Sweden	1						
Thailand	1						
Undefined	3						

Table 2: Articles included for qualitative synthesis by country, keyword, year and subject

The 49 articles prioritised for qualitative synthesis comprised 46 research articles and three reviews. The following analysis applies only to these articles. The majority of articles included for qualitative synthesis were published in NATO member states (See **TABLE 2**), with publications from the United States and United Kingdom featuring prominently in the results. These are complimented by publications from Australia and Canada, which may infer a positive correlation between member states of the "Five Eyes" signals intelligence (SIGINT) alliance [11] and policy concerns regarding cyber–warfare. This inference may be supported by the author affiliation of eight of the included articles with the defence apparatus of three of the "Five Eyes" member states[1].

Analysing the included articles by subject area, the majority of studies concern the social sciences, complimented by engineering, arts and humanities and computer science (See **TABLE 2**). Analysing the included articles by year, there are two deltas which bear scrutiny (See **TABLE 2**). The first occurs in 2012, where the number of publications more than tripled from the previous year, which may be correlated *a priori* with the unprecedented Stuxnet cyber–attack against Iran's nuclear program in 2010 [12] and the public release of the United States Department of Defense's 2010 Annual Report to Congress concerning Military and Security Developments Involving the People's Republic of China [13], which warned for the first time of the Chinese People's Liberation Army's capacity–building for offensive cyber operations. The second delta occurs in 2015 where the number of publications increased by eight from the previous year, which may be correlated *a priori* with the 2014 Sony Pictures Entertainment Hack [14], a prominent confidentiality–related cyber–attack formally attributed by the United States Federal Bureau of Investigation and Department of Justice to threat actors affiliated with the North Korean government's intelligence apparatus[2].

Keyword analysis of the included articles (See **TABLE 2**) includes prominent keywords that appeared in the keyword section of more than two articles. While many are expected, such as "cyber war" and "cyber attack", two keywords offer potential insight regarding the context of the included articles, which may inform the data synthesis. The first, "attribution", included in the keyword section of three of the included articles for synthesis, concerns the policy issue of attributing offensive cyber operations to actors [15]. Covert cyber operations, or state actors leveraging proxies to conduct offensive cyber operations, may contravene Rule 149 of Customary IHL [3], regarding the responsibility of a state for violations of IHL committed directly or via proxy. This is particularly salient given the acknowledged connections between several state SIGINT apparatuses and cyber threat actors. An exemplar of this issue is the identified connection between the United States National Security Agency's elite Tailored Access Operations unit [16] and the Equation Group [17], a cyber threat actor group classified as an advanced persistent threat [18] and linked to cyber–attacks against Iran, Russia and Syria among others, allegedly including the Stuxnet cyber–attack [12].

The second prominent keyword, "Tallinn Manual", included in the keyword section of five of the included articles for synthesis, refers to the Tallinn Manual on the International Law Applicable to Cyber Warfare [9]. Initially published by the NATO Cooperative Cyber Defence Centre of Excellence in 2013[3], the Tallinn manual is a non–binding academic study concerning the governance of cyber–warfare under IHL, authored by a broad church of representatives from academia, civil society and government. Given its seminal status and influence on policy, which has been subject to rigorous scrutiny by legal scholars and practitioners [20, 21, 22], it will undoubtedly inform the synthesis of this review.

[1] **Author's note:** funding sponsorship status is unclear.
[2] **Author's note:** this attribution has been publicly disputed by members of the cyber–security community.
[3] **Author's note:** a second edition of the Tallinn Manual was published in 2017 [19].

Appendix

1. BibTeX databases for each stage of the PRISMA workflow outlined in this protocol (See **SECTION 10**), including articles included for qualitative synthesis, may be found at: `https://github.com/20180626/systematic-review-protocol`.

2. Active learning for Systematic Reviews (ASReview)

   - Source code: `https://github.com/asreview/asreview`
   - Documentation: `https://asreview.readthedocs.io`
   - White paper: Rens van de Schoot, Jonathan de Bruin, Raoul Schram, Parisa Zahedi, Jan de Boer, Felix Weijdema, Bianca Kramer, Martijn Huijts, Maarten Hoogerwerf, Gerbrich Ferdinands et al. Asreview: open source software for efficient and transparent active learning for systematic reviews. *arXiv preprint arXiv:2006.12166*, 2020.

3. General Architecture for Text Engineering (GATE)

   - Source code: `https://github.com/GateNLP/gate-core`
   - Documentation: Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Cristian Ursu, Marin Dimitrov, Mike Dowman, Niraj Aswani, Ian Roberts, Yaoyong Li et al. *Developing Language Processing Components with GATE Version 5:(a User Guide)*. University of Sheffield, 2009.
   - White paper: Hamish Cunningham. Gate, a general architecture for text engineering. *Computers and the Humanities*, 36(2):223–254, 2002.

References

[1] Jean Pictet. *The Geneva Conventions of 12 August 1949: Geneva convention for the amelioration of the condition of the wounded and sick in armed forces in the field*, volume 1. International Committee of the Red Cross, 1952.

[2] George Patterson Manson. Cyberwar: the united states and china prepare for the next generation of conflict. *Comparative Strategy*, 30(2):121–133, 2011.

[3] Knut Dà[1]*rmann* and Baptiste Rolle. *Customary international humanitarian law*, volume 1. Cambridge University Press, 2005.

[4] Herbert Lin. Cyber conflict and international humanitarian law. *Int'l Rev. Red Cross*, 94:515, 2012.

[5] Christopher J Eberle. Just war and cyberwar. *Journal of Military Ethics*, 12(1):54–67, 2013.

[6] International Committee of the Red Cross. Protocol additional to the geneva conventions of 12 august 1949, and relating to the protection of victims of international armed conflicts (protocol i), 8 june 1977, 1987.

[7] Samuli Haataja. The 2007 cyber attacks against estonia and international law on the use of force: an informational approach. *Law, Innovation and Technology*, 9(2):159–189, 2017.

[8] Joe Burton. NATOâs cyber defence: strategic challenges and institutional adaptation. *Defence Studies*, 15(4):297–319, 2015.

[9] Michael N Schmitt. *Tallinn manual on the international law applicable to cyber warfare*. Cambridge University Press, 2013.

[10] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Int J Surg*, 8(5):336–341, 2010.

[11] Corey Pfluke. A history of the five eyes alliance: possibility for reform and additions. *Comparative Strategy*, 38(4):302–315, 2019.

[12] Ralph Langner. Stuxnet: dissecting a cyberwarfare weapon. *IEEE Security & Privacy*, 9(3):49–51, 2011.

[13] Robert M Gates. *Military and Security Developments Involving the People's Republic of China 2010: Annual Report to Congress*. Diane Publishing, 2010.

[14] Stephan Haggard and Jon R Lindsay. North korea and the sony hack: exporting instability through cyberspace. *Asia-Pacific Issues*, (117):1, 2015.

[15] Randall R Dipert. The ethics of cyberwarfare. *Journal of Military Ethics*, 9(4):384–410, 2010.

[16] Steven Loleski. From cold to cyber warriors: the origins and expansion of NSAâstailored access operations (tao) to shadow brokers. *Intelligence and National Security*, 34(1):112–128, 2019.

[17] Charl van Der Walt. The impact of nation-state hacking on commercial cyber–security. *Computer Fraud & Security*, 2017(4):5–10, 2017.

[18] Adel Alshamrani, Sowmya Myneni, Ankur Chowdhary and Dijiang Huang. A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities. *IEEE Communications Surveys & Tutorials*, 21(2):1851–1877, 2019.

[19] Michael N Schmitt. *Tallinn manual 2.0 on the international law applicable to cyber operations*. Cambridge University Press, 2017.

[20] Oliver Kessler and Wouter Werner. Expertise, uncertainty, and international law: a study of the tallinn manual on cyberwarfare. *LJIL*, 26:793, 2013.

[21] Dieter Fleck. Searching for international rules applicable to cyber warfare–a critical first assessment of the new tallinn manual. *Journal of Conflict and Security Law*, 18(2):331–351, 2013.

[22] Wolff Heintschel von Heinegg. The tallinn manual and international cyber security law. *Yearbook of international humanitarian law*, 15:3–18, 2012.

[23] Rens van de Schoot, Jonathan de Bruin, Raoul Schram, Parisa Zahedi, Jan de Boer, Felix Weijdema, Bianca Kramer, Martijn Huijts, Maarten Hoogerwerf, Gerbrich Ferdinands et al. Asreview: open source software for efficient and transparent active learning for systematic reviews. *arXiv preprint arXiv:2006.12166*, 2020.

[24] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, Valentin Tablan, Cristian Ursu, Marin Dimitrov, Mike Dowman, Niraj Aswani, Ian Roberts, Yaoyong Li et al. *Developing Language Processing Components with GATE Version 5:(a User Guide)*. University of Sheffield, 2009.

[25] Hamish Cunningham. Gate, a general architecture for text engineering. *Computers and the Humanities*, 36(2):223–254, 2002.